

KLASSENEINTEILUNG BEI STICHPROBEN DIE EINFLÜSSE AUF DIE PARAMETERSCHÄTZUNG

DISSERTATION
zur Erlangung des Grades
eines Doktors der
wirtschaftlichen Staatswissenschaften
(Dr. rer. pol.)
des Fachbereichs Rechts- und
Wirtschaftswissenschaften
der
Johannes Gutenberg-Universität Mainz

vorgelegt von
Diplom-Volkswirt
Franz-Ferdinand Kuhn

1984



KLASSEN EINTHEILUNG
BEI STICHPROBEN
DIE EINFLÜSSE AUF DIE
PARAMETERSCHÄTZUNG

Digitized by the Internet Archive
in 2026

KLASSENEINTEILUNG BEI STICHPROBEN DIE EINFLÜSSE AUF DIE PARAMETERSCHÄTZUNG

DISSERTATION
zur Erlangung des Grades
eines Doktors der
wirtschaftlichen Staatswissenschaften
(Dr. rer. pol.)
des Fachbereichs Rechts- und
Wirtschaftswissenschaften
der
Johannes Gutenberg-Universität Mainz

vorgelegt von
Diplom-Volkswirt
Franz-Ferdinand Kuhn

1984

KLASSENEINTEILUNG
BEI STICHPROBEN
DIE EINFLÜSSE AUF DIE
PARAMETERSCHÄTZUNG

Erstberichterstatter: Prof.Dr. Erich Härtter
Datum der Berichterstattung: 04.01.1984

Zweitberichterstatter: Prof.Dr. Peter M. Schulze
Datum der Berichterstattung: 18.01.1984

Tag der mündlichen Prüfung: 15.02.1984

Inhaltsverzeichnis

VORBEMERKUNGEN	1
I. PROBLEMSTELLUNG UND ABGRENZUNG DES THEMAS	3
1. Klasseneinteilung bei Er- hebungen	3
1.1. Definition der Klasseneinteilung	3
1.2. Gründe für eine Klasseneinteilung	7
1.2.1. Klasseneinteilung zum Zweck der Vereinfachung von Stichproben	7
1.2.2. Klasseneinteilung zum Zweck der Datenerfassung	7
1.2.2.1. Die Erfassung von Vermögensbe- ständen privater Haushalte im Rahmen der Einkommens- und Ver- brauchsstichprobe von 1969	8
1.2.2.2. Technisch vorgegebene Klassen- einteilung	11
1.3. Zusammenfassung	12
2. Theoretische Grundlagen	13
2.1. Standort der Klasseneinteilung bei statistischen Untersuchungen	15

2.2.	Klasseneinteilung bei Erhebungen - Allgemeine Darstellung und Be- zeichnungen	17
3.	Zusammenfassung	24
II. DIE EINFLÜSSE AUF DIE PARAMETERSCHÄTZUNG		25
1.	Allgemeine Schätztheorie	25
1.1.	Die Stichprobe	26
1.1.1.	Einstufige Auswahl - Die unein- geschränkte Zufallsauswahl	27
1.1.2.	Mehrstufige Auswahl	29
1.1.3.	Die Stichprobe als mehrdimen- sionaler Zufallsvektor	30
1.1.4.	Die Stichprobenfunktion und ihre Eigenschaft	32
1.2.	Definition der Schätzfunktion	35
1.2.1.	Erwartungstreue Schätzfunktion	36
1.2.2.	Konsistente Schätzfunktion	38
1.2.2.	Effiziente Schätzfunktion	39
1.3.	Zusammenfassung	42
2.	Das Modell der Klasseneinteilung	44
2.1.	Modellannahmen	44
2.2.	Die diskrete Verteilung der Variablen X und ihre Parameter	46

2.3.	Der Zusammenhang zwischen der Zufallsvariablen X und Z	48
3.	Der Einfluß der Klassenein- teilung auf den Erwartungswert $E(Z)$ und die Varianz $\text{Var}(Z)$ einer beliebigen stetigen Verteilung $F(z)$	50
3.1.	Die stetige Verteilung $F(z)$ und ihre Parameter $E(Z)$ und $\text{Var}(Z)$	50
3.2.	Die diskrete Verteilung $F(x)$ und ihre Parameter $E(X)$ und $\text{Var}(X)$	51
3.3.	Bedingte Verteilungen, bedingte Erwartungswerte und bedingte Varianzen	55
3.3.1.	Die bedingte Wahrscheinlich- keitsalgebra und die bedingte Verteilungsfunktion	55
3.3.2.	Eigenschaften der bedingten Erwartung	59
3.4.	Der Einfluß der Klassenein- teilung auf den Erwartungs- wert $E(Z)$	61
3.5.	Der Einfluß der Klassenein- teilung auf die Varianz $\text{Var}(Z)$	70
3.6.	Zusammenfassung der Ergebnisse	85

1.	Eigenschaften der Exponential- verteilung	90
2.	Eigenschaften der Laplace-Ver- teilung	92
3.	Klasseneinteilung bei expo- nentialverteilter Zufalls- variablen Z	93
4.	Klasseneinteilung bei Laplace- verteilter Zufallsvariablen Z	99
5.	Interpretation der quanti- tativen Ergebnisse	103
5.1.	Der Gesamtfehler des Er- wartungswertes	103
5.1.1.	Bei exponentialverteilter Zufallsvariablen Z	103
5.1.2.	Bei Laplace-verteilter Zu- fallsvariablen Z	104
5.2.	Der Gesamtfehler der Varianz	106
5.2.1.	Bei exponentialverteilter Zufallsvariablen Z	106
5.2.2.	Bei Laplace-verteilter Zufallsvariablen Z	108

IV. ZUSAMMENFASSUNG UND KRITISCHE WÜRDIGUNG	112
ANHANG	115
Beispiel I	116
Beispiel II	118
Beispiel III	120
Beispiel IV	122
Beispiel V	124
Beispiel VI	126
Beispiel VII	128
Tabellen (1) und (2)	130
Tabelle (3)	131

LITERATURVERZEICHNIS

BAMBERG, G. u. BAUR, F.
Statistik
München, Wien 1980.

BETHEA, R.M.
Statistical Methods for Engineers and Scientists
New York, Basel 1975.

BICKEL, P.J. u. DOKSUM, K.A.
Mathematical Statistics
San Francisco, Düsseldorf, Johannesburg 1977.

BLEYMÜLLER, J. u.a.
Statistik für Wirtschaftswissenschaftler
München 1979.

CHOU, YA - LUN
Statistical Analysis
London, New York, Sydney, Toronto 1969.

COCHRAN, W.G.
Stichprobenverfahren
Berlin 1972.

EULER, M.
Die Verteilung des Geldvermögens der privaten
Haushalte nach sozialen Gruppen
Allgemeines Statistisches Archiv, Bd. 57, 1976.

derselbe

Der Beitrag der Einkommens- und Verbrauchsstichprobe
zur Erfassung der Vermögensbestände privater Haushalte
Allgemeines Statistisches Archiv, Bd. 56, 1975.

FERSCHL, F.

Deskriptive Statistik
Würzburg, Wien 1980.

FISZ, M.

Wahrscheinlichkeitsrechnung und mathematische Statistik
Berlin 1970.

FREUND, J.E.

Mathematical Statistics
London 1972.

GÄNSSLER, P. u. STUTE, W.

Wahrscheinlichkeitstheorie
Berlin, Heidelberg, New York 1977.

HACKL, KATZENBEISSER, PANNY

Statistik
München 1981.

HAUSER, S.

Statistische Verfahren zur Datenbeschaffung
und Datenanalyse
Freiburg 1981.

HAYS, W.L. u. WINKLER, R.L.

Statistics: Probability, Inference and Decision
New York, Chicago, San Francisco, ... 1971.

HELLER, K., ROSEMAN, B.

Planung und Auswertung empirischer Untersuchungen
Stuttgart 1974.

HOCHSTÄDTER, D.

Einführung in die statistische Methodenlehre
Frankfurt/Main 1978,

KELLERER, H.

Theorie und Technik des Stichprobenverfahrens
Einzelschriften der Deutschen Statistischen Gesellschaft
NR. 5, München 1974.

KREYSZIG, E.

Statistische Methoden und ihre Anwendungen
Göttingen 1973.

MENGES - SKALA

Statistik 2, Daten
Opladen 1976.

MORONEY, M.J.

Einführung in die Statistik - Teil 1
München, Wien 1980.



RENYI; A.

Wahrscheinlichkeitsrechnung

Berlin 1962.

RINNE, H., ICKLER, G.

Statistik Teil 1

München 1979.

RINNE, H.

Statistik Teil 3

München 1979.

SACHS, L.

Angewandte Statistik

Berlin, Heidelberg, New York 1978.

SCHNEEWEISS, H.

Modelle mit Fehlern in den Variablen

Methods of Operations Research. Bd. 37, 1980.

derselbe

Ökonometrie

Würzburg, Wien 1978.

SPIEGEL; M.R.

Theory and Problems of Statistics

New York 1972.

STANGE, K.

Angewandte Statistik erster Teil
Berlin, Heidelberg, New York 1970.

derselbe

Angewandte Statistik zweiter Teil
Berlin, Heidelberg, New York 1971.

derselbe

Bayes-Verfahren; Schätz- und Testverfahren bei
Berücksichtigung von Vorinformationen
Berlin, Heidelberg, New York 1977.

STATISTISCHES BUNDESAMT (Hersg.)

Probleme der Erfassung von Vermögenswerten
privater Haushalte im Rahmen der Einkommens-
und Verbrauchsstichprobe 1969
Wista, 22 Jahrg., Heft 12, Dez. 1970.

dasselbe

Stichproben in der amtlichen Statistik
Stuttgart, Mainz 1960.

STATISTISCHES LANDESAMT SCHLESWIG - HOLSTEIN
Volkszählung 1983, Wohnungsbogen.

STURGES, H.A.

The Chioce of Class Interval
Journal of American Statistical Association, March 1929.

VOGEL; W.
Wahrscheinlichkeitstheorie
Göttingen 1970.

WAGENFUHR, R.
Statistik leicht gemacht, Bd. 1
Köln 1971.

WONNACOTT; T.H., WONNACOTT, R.J.
Introductory Statistics for Business and Economics
New York, London, Sydney, Toronto 1972.

ZACKS, S.
The Theory of Statistical Inference
New York, London, Sydney, Toronto 1971.

VORBEMERKUNGEN

Wieso beschäftigt sich ausgerechnet ein Wirtschafts-
wissenschaftler mit dem Thema "Klasseneinteilung
bei Stichproben - Die Einflüsse auf die Parameter-
schätzung"?

Im Rahmen der Wirtschaftswissenschaften dienen die
statistischen Methoden vor allem zur Stützung von
ökonomischen Hypothesen oder zur Findung von Zusammen-
hängen ökonomischer Größen. Diese gehen dann mit in
den wirtschaftspolitischen Entscheidungsprozeß ein.

Ausgangspunkt einer jeden statistischen Untersuchung
ist die Erfassung von Daten. Aber gerade bei ökono-
mischen Fragestellungen bereitet es oft Schwierigkeiten,
an die entsprechenden Daten zu gelangen. Kann man näm-
lich nicht auf Sekundärstatistiken (Statistische Jahr-
bücher, Melderegister etc.) zurückgreifen, bleibt nur
der Weg über Befragungen. Wie die Diskussion über die
geplante Volkszählung im April dieses Jahres zeigt,
stößt man hier zunehmend auf Widerstand bei den Befragten.
Sie fürchten um ihre Anonymität. Die Praxis hilft sich,
indem sie nur nach Größenordnungen fragt. Im Fragebogen
des Statistischen Landesamtes Schleswig-Holstein zur
geplanten Volkszählung 1983 findet man zum Beispiel
die Frage:

"Wann ist der Haushalt in diese Wohnung eingezogen?

vor 1970
1970 - 1974
1975 - 1980
1981
1982
1983 " 1).

1) Statistisches Landesamt Schleswig Holstein,
Volkszählung 1983, Wohnungsbogen, Frage 4.

Hier werden den Befragten Größenklassen vorgegeben, in die sie sich einstufen. ¹⁾ Aber auch die einzelnen Jahre stellen für sich eine Klasse dar, denn jedes läuft vom 1. Januar bis zum 31. Dezember, also 365 Tage. Gerade für den Ökonomen, der auf der Basis einer statistischen Untersuchung wirtschaftspolitische Entscheidungen fällt oder Zusammenhänge von ökonomischen Größen herausarbeiten will, ist es es wichtig, die Wirkungen dieser Art von Datenerfassung zu kennen.

Damit man die Einflüsse der Klasseneinteilung bei Stichproben untersuchen kann, wird folgendes vorausgesetzt : Das Datenmaterial ist skaliert. Die Skala der Daten bildet die reelle Zahlenachse (zum Beispiel: D-Mark-Beträge, Jahreszahlen usw.).

1) Vgl.: Hierzu auch die Ausführungen des Teil I, Abschnitt 1.2.2.1.

I. PROBLEMSTELLUNG UND ABGRENZUNG DES THEMAS

1. Klasseneinteilung bei Erhebungen

1.1. Definition der 'Klasseneinteilung

Es liegt eine Erhebung vor. Für die Klasseneinteilung spielt es im Prinzip keine Rolle, ob etwas total oder durch eine Stichprobe erhoben wird. Die theoretische Erörterung geht jedoch meist von einer Stichprobe aus.

Oftmals treten in einer solchen Stichprobe sehr viele zahlenmäßig verschiedene Werte auf. In diesem Fall ist es sehr schwer, einen Überblick darüber zu erhalten, wie sich die einzelnen Merkmalsausprägungen verteilen; das heißt festzustellen, welche Werte häufiger und welche seltener erscheinen. Graphisch veranschaulicht ergibt eine solche Stichprobe eine unübersichtliche Vielzahl von Punkten, da nun einmal fast alle Ausprägungen verschieden sind.

Um dennoch Aufschluß über die Verteilung zu erlangen, - zum Beispiel durch Zeichnung von Histogrammen, Stabdiagrammen, Kreisdiagrammen oder Summenhäufigkeitesverteilungen - empfehlen statistische Lehrbücher eine Klasseneinteilung. Durch sie vereinfacht sich zum einen die Stichprobe. Zudem kann sie graphisch besser dargestellt werden. Und drittens verringert sich der Rechenaufwand der weiteren Analyse.

Bei eindimensionalen Problemen geht man von dem Intervall aus, in dem alle Stichprobenwerte liegen (dem Streubereich). Dieses wird in Teilintervalle, die sogenannten Klassen, aufgeteilt. Die Anzahl der Elemente

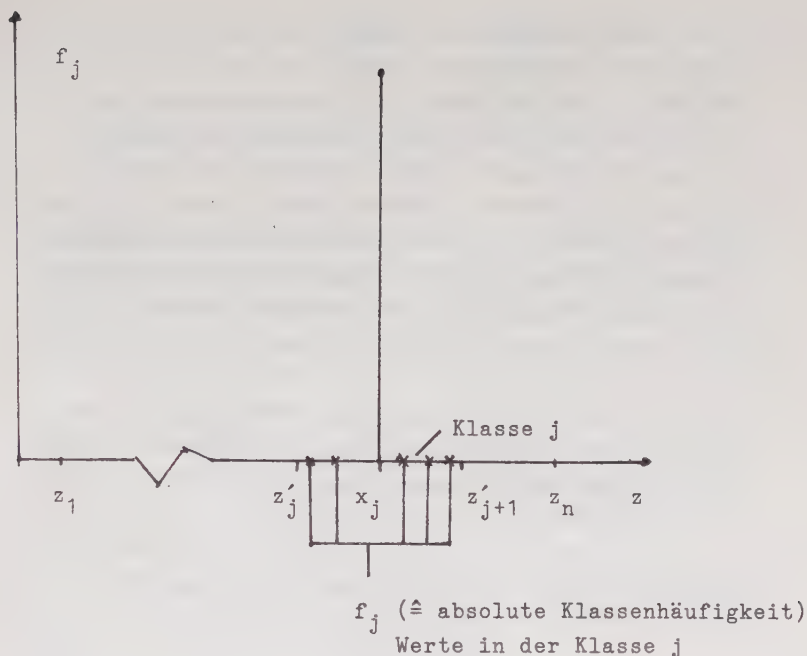
in einer jeden Klasse heißen absolute Häufigkeiten oder besser absolute Klassenhäufigkeiten, die Intervalllängen Klassenbreiten.

Bei mehrdimensionalen, etwa r -dimensionalen Problemen, erscheinen die Stichprobenwerte als Punkte in einem entsprechenden Raum $\mathbb{R}^r$. Die Klassenbildung vollzieht sich dann parallel zu jeder Koordinatenachse dieses Raumes. Es entstehen Teilräume, die wiederum als Klassen bezeichnet werden. Die Anzahl der Punkte in einem solchen Teilraum (Klasse) bildet auch hier wieder die absolute Klassenhäufigkeit.

Fazit: Bei klassifiziertem Datenmaterial treten die ursprünglichen Werte nicht mehr in Erscheinung, sondern nur noch die einzelnen Teilintervalle mit ihren absoluten Häufigkeiten.

Zur weiteren Verarbeitung der Stichprobe nimmt man in der Literatur an, daß alle Werte in einer jeden Klasse in den zugehörigen Klassenmitten (Mittelpunkte der Klassen) liegen. Sie sind damit Repräsentanten für die Werte innerhalb einer jeden Klasse.

Die folgende Abbildung soll die Klasseneinteilung für ein eindimensionales Problem verdeutlichen.



Die Originalwerte $z_1, z_2, \dots, z_n$ seien der Größe nach geordnet. Das Intervall zwischen dem kleinsten Wert z_1 und dem größten z_n , in dem alle Elemente der Stichprobe liegen (der Streubereich), wird in k paarweise punktfremde Teilintervalle $[z'_j; z'_{j+1})$ ($j = 1, 2, \dots, k$) aufgeteilt. Die jeweilige Klassenmitte x_j repräsentiert alle Werte in ihrer Klasse j . Die einzelnen z -Werte im Intervall j - in der Abbildung durch Kreuze gekennzeichnet - werden somit bei der weiteren Analyse nicht mehr berücksichtigt.

Auf die Information, die in der Verteilung der Stichprobenwerte innerhalb einer jeden Klasse steckt, verzichtet man also. Durch die Gruppenbildung entsteht ein Informationsverlust.

Für ein zweidimensionales Problem mit den Variablen z_1 und z_2 zum Beispiel erscheinen die Stichprobenwerte als Punkte in einem z_1/z_2 Koordinatensystem. Zur Klassenbildung unterteilt man das Intervall der z_1 -Achse, in dem alle z_1 -Werte liegen, den Streubereich, in Teilintervalle. Ebenso wird mit der z_2 -Achse verfahren. Damit entstehen Teilflächen - die Klassen - in der z_1/z_2 -Ebene. Die "Mittelpunkte" dieser Klassen bilden die Klassenmitten. Sie repräsentieren alle Punkte in den jeweiligen Teilflächen.¹⁾

Auch hier verursacht die Gruppenbildung (Klassenverteilung) einen Informationsverlust. Ziel dieser Arbeit ist es, diesen Informationsverlust zu untersuchen und zu quantifizieren.

Doch zunächst soll auf die Klasseneinteilung bei Erhebungen noch näher eingegangen werden.

-
- 1) Zur Klasseneinteilung vgl. u.a. :
Bley Müller, J. u.a.: Statistik für Wirtschaftswissenschaftler, München 1979, S. 9 ff.
Fersch, F. : Deskriptive Statistik, Würzburg-Wien 1980, S. 97 ff.
Hackl, Katzenbeisser, Panny : Statistik, München 1981, A. 11 ff.
Hauser, S. : Statistische Verfahren zur Datenbeschaffung und Datenanalyse, Freiburg 1981, S. 20 ff.
Heller, K.; Rosemann, B. : Planung und Auswertung empirischer Untersuchungen, Stuttgart 1974, S. 95 ff.
Kreyszig, E. : Statistische Methoden und Ihre Anwendungen, Göttingen 1973, S. 29 ff.
Moroney, M.J. : Einführung in die Statistik - Teil 1, München-Wien 1980, S. 77 ff.
Rinne, H.; Ickler, G. : Statistik Teil 1, München 1979 S. 63 ff.
Stange, K. : Angewandte Statistik erster Teil, Berlin-Heidelberg-New York 1970, S. 9 ff.
Derselbe : Angewandte Statistik zweiter Teil, Berlin-Heidelberg-New York 1971, S. 2 ff.

1.2. Gründe für eine Klasseneinteilung

1.2.1. Klasseneinteilung zum Zweck der Vereinfachung von Stichproben

Die Definition der Klasseneinteilung verdeutlicht an sich schon den bekanntesten Grund für eine Gruppenbildung. Er soll hier der Vollständigkeit halber noch einmal beschrieben werden.

Gegeben ist eine Erhebung in Form einer Stichprobe oder einer Totalerhebung. Die einzelnen Daten weisen sehr viele zahlenmäßig verschiedene Werte auf. Hier führt man eine Klasseneinteilung nach der im ersten Abschnitt beschriebenen Vorgehensweise durch. Die Stichprobe vereinfacht sich, sie läßt sich graphisch leichter darstellen, und der Rechenaufwand der weiteren Analyse wird geringer. Daraus folgt

Erster Grund :

Klasseneinteilung dient zur Vereinfachung von Stichproben nach der Erhebung.

1.2.2. Klasseneinteilung zum Zweck der Datenerfassung

Um zu zeigen, wie eine Klasseneinteilung in der Praxis zur Datenerfassung eingesetzt wird, soll zuerst ein Beispiel aus der ökonomischen Anwendung beschrieben werden.

1.2.2.1. Die Erfassung von Vermögensbeständen privater Haushalte im Rahmen der Einkommens- und Verbrauchsstichprobe von 1969

Im Rahmen der zweiten Einkommens- und Verbrauchsstichprobe 1969 gehörte die Erfassung der Vermögensbestände privater Haushalte zum erstenmal mit zum gesetzlichen Auftrag. Ziel war es, auf statistischem Weg etwas Licht in das Dunkel der Vermögensverteilung - insbesondere der Verteilung auf Bevölkerungsgruppen - zu bringen, die zu den brennenden sozial- und gesellschaftspolitischen Problemen gehört. Die hierbei entstehenden Schwierigkeiten der Begriffsabgrenzung, der Bewertung und der Zuordnung interessieren im Rahmen dieser Arbeit nicht. Sie sind aber in der Literatur ausführlich behandelt.¹⁾

In den Schlußinterviews, die Anfang 1970 stattfanden, wurde nach folgenden Vermögensbeständen gefragt:

- Guthaben auf Sparkonten
- Guthaben auf Bausparkonten
- Lebensversicherungssummen
- Haus- und Grundbesitz.

Gerade bei der Offenlegung von Einkommens- und Vermögensverhältnissen kommt es jedoch durch psychologische Hemmnisse zu Erhebungsausfällen. Um diese zu vermeiden, wurde darauf verzichtet, die genauen Bestände der jeweiligen Vermögenswerte zu erfassen. Man bildete verschiedene Größenklassen und bat die Haushalte, sich

1) Vgl.: Statistisches Bundesamt (Hersg.): Probleme der Erfassung von Vermögenswerten privater Haushalte im Rahmen der Einkommens- und Verbrauchsstichprobe 1969; Wista, 22 Jahrg., Heft 12, Dez. 1970, S. 601 ff.

anhand ihrer Unterlagen in diese einzustufen.¹⁾

Zur Erfassung der Spar-, Bausparguthaben und der Lebensversicherungssummen sollten die Haushalte die jeweiligen Guthaben addieren und sich in die entsprechenden Bestandsgruppen einordnen. Ebenso stellte man die Einheitswerte aller privaten Grundstücke und Gebäude von 1965 fest. Die Erhebung der Wertpapierbestände sollte zum Tageswert erfolgen. Da die Haushalte oft verschiedenartige Geldanlagen halten, ist eine solche Vorgehensweise mit hohem Arbeitsaufwand verbunden. Sie müßten jedes einzelne Wertpapier zum Tageskurs bewerten. Durch die Vorgabe von nur sechs Größenklassen erleichterte man ihnen die Arbeit.²⁾

Wie dieser Fall aus der ökonomischen Praxis zeigt, benutzt man eine Klasseneinteilung auch zur Datenerfassung. Zum Beispiel dann, wenn wir auf Primärerhebungen - genauer auf mündliche oder schriftliche Befragungen - angewiesen sind und psychologische Hemmnisse abbauen oder den Arbeitsaufwand der Datenerfassung vereinfachen müssen. In diesen Fällen ist dann die Klasseneinteilung a priori vorgegeben.

In den Anwendungen kommt dies oft bei Forschungen auf den Gebieten der Markt- und Absatzanalyse oder bei ökonomischen Untersuchungen wie in dem oben beschriebenen Beispiel vor.

1) Vgl.: Statistisches Bundesamt (Hersg.): Probleme der Erfassung von ..., a.a.O., S. 602.

2) Vgl.: Euler, M.: Die Verteilung des Geld- und Grundvermögens der privaten Haushalte nach sozialen Gruppen; in: Allgemeines Statistisches Archiv, Bd. 57, 1976, S. 225 ff. und derselbe: Der Beitrag der Einkommens- und Verbrauchsstichproben zur Erfassung der Vermögensbestände privater Haushalte; in: Allgemeines Statistisches Archiv Bd. 56, 1979, S. 99 ff.

Damit ergeben sich die Gründe zwei und drei.

Zweiter Grund:

Ist man auf Primärerhebungen angewiesen und tauchen bei der Offenlegung von Daten psychologische Hemmnisse auf, so faßt die Praxis einzelne Merkmalsausprägungen in Gruppen - Größenklassen - zusammen. Die Befragten stufen sich dann anhand ihrer tatsächlichen Gegebenheiten in eine dieser Größenklassen ein. Hierdurch läßt sich höhere Anonymität der Befragten erreichen, psychologische Hemmnisse werden abgebaut, und der Schutz der persönlichen Daten bleibt eher gewährleistet.

Gerade dieser Punkt gewinnt bei fortschreitendem Informationsbedarf der Entscheidungsträger in Politik und Wirtschaft sowie gleichzeitiger Forderung nach mehr Schutz der persönlichen Daten immer weiter an Bedeutung. Insbesondere dann, wenn man bei ökonomischen Untersuchungen auf private Datenträger angewiesen ist.

Dritter Grund:

Bei Erhebungen, bei denen die Feststellung der Daten sehr arbeitsaufwendig ist, kann man durch Bildung von Gruppen - wie das Beispiel zeigt - ihre Erfassung vereinfachen.

1.2.2.2. Technisch vorgegebene Klasseneinteilung

Es sollen technische Größen wie Längen, Rauminhalte, Gewichte usw. erhoben werden. Hierzu benötigt man ein Meßgerät, das nur eine technisch vorgegebene Genauigkeit aufweist. Die damit erhaltenen Werte sind gerundet. Zum Beispiel: Durch eine bestimmte Waage kann das Gewicht nur auf Gramm genau festgestellt werden. Milligramm oder noch kleinere Einheiten sind nicht meßbar. Diese Rundung von Daten durch die technische Ausstattung von Meßgeräten stellt eine Klasseneinteilung dar.

Bei der Verwendung von ökonomischen Größen - zum Beispiel aus dem Statistischen Jahrbuch - stellt man meist fest, daß sie nur in gerundeten DM-Beträgen vorliegen. Die genauen Werte sind hier nicht verfügbar. Damit handelt es sich auch in diesem Fall um klassifizierte Daten. Auch hier ist die Klasseneinteilung a priori gegeben.

Vierter Grund:

Klassifizierte Daten entstehen auf Grund von technischen Gegebenheiten wie die Genauigkeit von Meßgeräten oder von Veröffentlichungen der amtlichen Statistik.

1.3. Zusammenfassung

Es zeigt sich: Dient die Gruppenbildung der Vereinfachung von Stichproben (erster Grund), so stehen sowohl klassifizierte als auch Originaldaten bereit. Die Klasseneinteilung wurde auf Basis einer bestimmten Erhebung mit wahren Werten durchgeführt. Man kann demnach nicht nur die klassifizierten Daten analysieren sondern auch auf die Originaldaten zurückgreifen, wenn man sie jederzeit zur Verfügung hält. Eventuelle Fehler ließen sich so vermeiden. Dazu müßte die Stichprobe aber doppelt abgespeichert werden. Eine weiter fortschreitende Technologie der EDV ermöglicht das ohne Schwierigkeiten. Es stellt sich nun die Frage: Lohnt sich diese Arbeit? - Oder: Wann ist das Risiko eines Fehlers durch die Verwendung von klassifizierten Daten größer als der Aufwand der zweimaligen Abspeicherung?

Wird hingegen eine Klasseneinteilung zur Datenerfassung vorgenommen (zweiter bis vierter Grund) (Klasseneinteilung ist a priori vorgegeben), so treten die Originaldaten nicht mehr in Erscheinung. Die statistische Untersuchung kann sich nur auf Intervalle mit ihren absoluten Häufigkeiten stützen. Die Problematik liegt auf der Hand: Sind die aus klassifizierten Daten gewonnenen Schlüsse brauchbar, oder beeinflußt die Gruppenbildung die Ergebnisse?

Für die Beantwortung beider Fragenkomplexe tut es not, den Informationsverlust zu quantifizieren. Dann läßt sich auch analysieren: Wie genau muß man Daten erfassen, wenn eine vorgegebene Sicherheit des statistischen Ergebnisses erreicht werden soll?

2. Theoretische Grundlagen

2.1. Standort der Klasseneinteilung bei statistischen Untersuchungen

Wie schon angedeutet, soll der Informationsverlust, der durch eine Klasseneinteilung entsteht, untersucht und quantifiziert werden. Hierzu ist es notwendig zu wissen, an welcher Stelle einer statistischen Analyse die Klassenbildung eigentlich eingreift.

Gegeben sei ein Phänomen, über das man Aufschluß erhalten will. Wir untersuchen es entweder gesamt (Totalerhebung) oder teilweise (Stichprobe). Durch die Erhebung und Messung seiner Realisationen entstehen Daten. Sie werden verarbeitet und danach einem Entscheidungsträger als Informationen übermittelt. Er transformiert sie in Aktionen. Solche Aktionen können betriebliche oder politische Maßnahmen oder auch statistische Urteile im Sinne einer Schätzung, einer Hypothesenüberprüfung oder einer Prognose sein.¹⁾

Der Begriff der Daten steht damit im Mittelpunkt einer statistischen Untersuchung. Menges definiert sie wie folgt: "Statistische Daten sind Symbole, welche aus einer Erhebung der Realisationen von Phänomenen und deren Messung hervorgehen und durch geeignete Verarbeitung zu Informationen werden, welche ihrerseits Grundlage für statistische Entscheidungen bilden"²⁾.

1) Vgl.: Menges-Skala: Statistik 2, Daten, Opladen 1976; S. 24 - 25.

2) Ebenda, S. 32.

Die Transformation von Daten in statistische Information vollzieht sich nun in mehreren Stufen.

In der ersten Stufe nimmt man an, daß die Daten zufällige Realisationen von Zufallsvariablen sind. Über sie wird ein bestimmtes stochastisches Modell gestülpt. Es soll möglichst genau die wahren Eigenschaften des Phänomens widerspiegeln.

Daran schließt sich die zweite Stufe an. Die Daten werden in Maßzahlen transformiert. Man berechnet zum Beispiel Mittelwerte und Streuungsmaße.

Die Ergebnisse der zweiten Stufe werden in einer dritten zu einem statistischen Resultat verarbeitet. Dieses stellt entweder eine Schätzung, die Präferenz einer Hypothese oder einer Prognose dar.¹⁾

Um festzustellen bei welcher Stufe einer statistischen Untersuchung die Klassenbildung eingreift, müssen wir zwei Fälle unterscheiden:

- a) Sollen Befragungen durchgeführt werden, und muß man dazu psychologische Hemmnisse abbauen oder ist der Arbeitsaufwand der Datenerfassung zu groß, so bildet die Praxis a priori Größenklassen. Die Erhebung besteht damit nur aus diesen Intervallen mit ihren absoluten Häufigkeiten. Das gleiche gilt, wenn aus Veröffentlichungen der amtlichen Statistik nur gerundete Daten entnommen werden können oder Meßgeräte nur eine bestimmte Genauigkeit aufweisen.

1) Vgl.: Menges-Skala: Statistik 2, ..., a.a.O., S. 44 - 45.

- b) Man nimmt eine Klassenbildung vor, um zum Beispiel Stichproben nach der Erhebung zu vereinfachen. Hier stehen sowohl Originaldaten als auch klassifizierte Daten zur weiteren Verarbeitung zur Verfügung.

Im Fall a) setzt die Klasseneinteilung schon bei der Messung und Erhebung eines Phänomens ein. Hier treten die einzelnen Realisationen nicht in Erscheinung. Für die weitere statistische Untersuchung stehen nur noch Intervalle mit ihren absoluten Häufigkeiten zur Verfügung. Das heißt, die erste Stufe der Verarbeitung der Daten lautet in diesem Fall: Die Klassen bzw. die Klassenmitten sind zufällige Realisationen einer Zufallsvariablen.

Führt man eine Klasseneinteilung durch, um zum Beispiel Stichproben zu vereinfachen - Fall b) -, setzt die Klassenbildung in der zweiten Stufe ein. Transformiert man diese dann in Datenmaße und schließt sich daran die dritte Stufe an, muß für die erste wieder angenommen werden: Die Klassen bzw. Klassenmitten sind zufällige Realisationen einer Zufallsvariablen. Damit kann Fall a) wie b) behandelt werden, wenn man - wovon wir ausgehen wollen - die Originaldaten nicht weiter berücksichtigt.

Zur Bestimmung des Informationsverlusts, der durch die Klasseneinteilung entsteht, geht man von folgenden Annahmen aus: Es soll ein Phänomen untersucht werden, das einen Zufallscharakter hat. Die Daten entstehen durch Messung von Realisationen dieser Zufallsvariablen. Die einzelnen Daten lassen sich aus den genannten Gründen nicht erfassen. Es sind nur Größenklassen beobachtbar, in denen die wahren Werte liegen. Diese bewegen sich innerhalb den vorgegebenen Intervallgrenzen.

Die Schwankungen hängen einmal von der Klassenbreite, zum anderen von der Form des unterstellten Zufallsprozesses und der jeweiligen Realisation ab.

Die Klasseneinteilung darf allerdings nicht mit den statistischen Fehlern verwechselt werden. Sie werden im allgemeinen in zufällige - (random error) und systematische Fehler (bias) aufgeteilt.

Zufällige Fehler entstehen zum Beispiel durch ungenaues Ablesen eines Meßgerätes; das heißt, sie sind zufallsabhängig. Häufig wird hier angenommen, daß ihr Erwartungswert null ist und sie unabhängig von der Realisation des Meßwertes auftreten. Der zufällige Fehler hat damit sogar ein eigenes stochastisches Modell.

Systematische Fehler sind dadurch gekennzeichnet, daß jeder Wert um eine bestimmte Größe verändert ausgewiesen wird.

Beide haben somit gemeinsam: Sie treten unabhängig von dem gerade beobachteten Wert auf.¹⁾ Bei Klasseneinteilung hingegen - dies sei hier noch einmal hervorgehoben und wird in dem später verwendeten Modell verdeutlicht - sind die Schwankungen in einem bestimmten Intervall sehr stark abhängig von dem unterstellten Zufallsprozeß des zu untersuchenden Phänomens und der beobachteten Gruppe (Klasse).

Um den Informationsverlust zu quantifizieren und seine Einflußgrößen zu bestimmen, wollen wir folgende

1) Vgl.: Menges-Skala: Statistik 2, a.a.O., S. 273.
Schneeweiß, H.: Ökonometrie, Würzburg-Wien 1978, S. 219 und derselbe: Modelle mit Fehlern in den Variablen, in: 'Methods of Operations Research', Bd. 37, S. 47.

Annahme treffen: Statistische Fehler in der oben beschriebenen Form treten nicht auf. Das heißt, wird ein Intervall beobachtet, so liegt der tatsächliche Wert auch in diesem Intervall.

2.2. Klasseneinteilung bei Erhebungen - Allgemeine Darstellung und Bezeichnungen

Bei der folgenden Darstellung und Analyse des genannten Problems möchte ich mich auf den eindimensionalen Fall beschränken, um die Einflußgrößen des Informationsverlusts herauszuarbeiten. Zwar spielen in den ökonomischen Anwendungen mehrdimensionale Probleme die weit wichtigere Rolle, aber auch unter dieser vereinfachten Annahme wird die Problematik deutlich. Zudem lassen sich die meisten Ergebnisse ohne Schwierigkeiten auf den komplizierteren Fall übertragen.

Zu untersuchen ist nun ein Phänomen, dessen Merkmal Z eine Zufallsvariable mit der stetigen Dichte $f(z)$ darstellt. Der Erwartungswert $E(Z) = \mu_z$ und die Varianz $\text{Var}(Z) = \sigma_z^2$ von Z sollen existieren.

Die Dichte $f(z)$ hat die Eigenschaften

$$\int_{-\infty}^{\infty} f(z) \, dz = 1 \quad \text{und}$$

$$f(z) \geq 0 \quad \text{für } z \text{ aus } \mathbb{R}.$$

Der Erwartungswert ist dann

$$E(Z) = \int_{-\infty}^{\infty} z f(z) dz = \mu_z$$

und die Varianz

$$\text{Var}(Z) = \int_{-\infty}^{\infty} [z - E(Z)]^2 f(z) dz = \sigma_z^2$$

Die Dichte $f(z)$ bzw. die Verteilungsfunktion $F(z)$ werden als gegeben angenommen. Die wahren Werte der Parameter μ_z und σ_z^2 sind unbekannt. Sie will man zum Beispiel durch eine Stichprobe vom Umfang n schätzen.

Aus den genannten Gründen sind die tatsächlichen Realisationen $z_1, z_2, \dots, z_1, \dots, z_n$ nicht beobachtbar, sondern nur Größenklassen. Es soll also eine Intervalleinteilung vorgenommen werden. Hierzu unterteilt man die z -Achse in abzählbar unendlich viele Teilintervalle (Klassen)

$$[z'_j; z'_{j+1}) \quad j = \dots, -2, -1, 0, 1, 2, \dots$$

Die zugehörigen Klassenmitten berechnen sich aus

$$x_j = \frac{z'_j + z'_{j+1}}{2},$$

und ihre Breiten durch

$$2c_j = z'_{j+1} - z'_j \quad .$$

Tritt nun in der Stichprobe der nicht beobachtbare Wert z_i auf, für den gilt

$$z'_j \leq z_i < z'_{j+1} \quad ,$$

so ordnet man ihm den Wert x_j zu.

Damit erscheinen die einzelnen x_j mit den absoluten Häufigkeiten f_j ($f_j \hat{=}$ Anzahl der Elemente in der Klasse j).

Bei einer speziellen Stichprobe vom Umfang n sind mehrere der abzählbar unendlich vielen Klassen unbesetzt. Es treten also Intervalle auf, deren absolute Häufigkeit $f_j = 0$ sind. Nur für endlich viele gilt $f_j \neq 0$.

Damit ergeben sich der Stichprobenumfang wieder aus

$$n = \sum_{-\infty}^{\infty} f_j$$

und die relativen Klassenhäufigkeiten durch

$$\hat{p}_j = f_j / n \quad .$$

Aus dieser klassifizierten Stichprobe lassen sich nun die Schätzwerte für μ_z und σ_z^2 berechnen:

Man hat

$$\hat{\mu}_x = 1/n \sum_{-\infty}^{\infty} x_j f_j = \bar{x} \quad \text{für } \mu_z,$$

und

$$\hat{\sigma}_x^2 = 1/n \sum_{-\infty}^{\infty} f_j (x_j - \bar{x})^2 = s_x^2 \quad \text{für } \sigma_z^2.$$

Wenn man die ursprünglichen Realisationen z_i ($i = 1, 2, \dots, n$) zur Verfügung hätte, könnte man

$$\hat{\mu}_z = 1/n \sum_{i=1}^n z_i = \bar{z} \quad \text{für } \mu_z$$

sowie

$$\hat{\sigma}_z^2 = 1/n \sum_{i=1}^n (z_i - \bar{z})^2 = s_z^2 \quad \text{für } \sigma_z^2$$

benutzen.

Nur wenn zutrifft, daß alle Realisationen von Z in den zugehörigen Klassenmitten x_j liegen, ist
- von günstigen Sonderfällen abgesehen -

$$\bar{x} = \bar{z}$$

und

$$s_x^2 = s_z^2 \quad .$$

Die Literatur nimmt die obige Annahme oft an, damit sie die Schätzwerte aus den klassifizierten Daten für die weitere Analyse benutzen kann, und empfiehlt die Klassenbildung, wie folgt vorzunehmen :

- Die Klassenbreiten $2c_j$ sollen möglichst konstant $2c$ sein.
- Die Anzahl der Klassen muß die verfügbaren Werte z_1 , daß heißt den Stichprobenumfang und den Streubereich berücksichtigen.

Es sollen also weder zu viele noch zu wenige Klassen gebildet werden. Teilt man die Stichprobe in zu viele Intervalle auf, so bleibt sie unübersichtlich; das Ziel der Klassenbildung wird nicht erreicht. Bildet man hingegen zu wenige, so kann die empirische Verteilung verfälscht werden. Nur durch solche Überlegungen kommen Faustregeln zustande, die die Wahl der Anzahl der Klassen oder die Intervalllängen bei Konstanz der Breiten bestimmen. So zum Beispiel:

Man soll nicht mehr als $\sqrt{n}$
($n \hat{=}$ Stichprobenumfang) Klassen wählen.¹⁾

1) Vgl.: Fersch, F.: Deskriptive Statistik, a.a.O., S. 29 und Hochstädter, D.: Einführung in die statistische Methodenlehre, Frankfurt/Main 1978, S. 35.

Oder die Formel von Sturges:¹⁾

Die Wahl der Klassenbreite $2c$ soll
die Gleichung

$$2c = \frac{\text{Spannweite der Stichprobe}^{2)}}{1 + 3,322 \log n}$$

erfüllen.

Wobei: $2c \hat{=}$ Klassenbreite

$n \hat{=}$ Stichprobenumfang

Diese Vorgehensweise will die Stichprobe möglichst vereinfachen und gleichzeitig gewährleisten, daß die obige Voraussetzung ihre Gültigkeit behält.³⁾

Dient die Klassenbildung zur Datenerfassung, so kann man nicht nach diesen Regeln vorgehen. Denn von außen vorgegebene Gegebenheiten wie Datenschutz, die Genauigkeit von Meßgeräten oder der Arbeitsaufwand der Erfassung beeinflussen die Wahl der Anzahl der Intervalle und damit ihre Breite.

Damit muß der Unterschied zwischen $\bar{x}$ und $\bar{z}$ bzw. zwischen $\hat{s}_x^2$ sowie $\hat{s}_z^2$ bei der weiteren statistischen Analyse mit berücksichtigt werden. Nur Sheppard schätzt den Fehler durch die Klasseneinteilung unter folgenden Voraussetzungen ab:

-
- 1) Vgl.: Sturges, H.A.: The Choice of Class Interval, in: Journal of the American Statistical Association, March 1929.
 - 2) Vgl.: Wagenführ, R.: Statistik leicht gemacht, Bd. 1, Köln 1971, S. 65.
 - 3) Vgl.: hierzu auch: Stange, K.: Angewandte Statistik erster Teil, a.a.O., S. 9 ff.

- Die Zufallsvariable Z hat die stetige Dichte $f(z)$.
- Die Dichte $f(z)$ ist symmetrisch zu dem Erwartungswert μ_z von Z .
- Die Dichte $f(z)$ sei auf dem endlichen Intervall (a,b) beschränkt.
- Das Intervall (a,b) wird in k (endlich) Klassen eingeteilt mit der konstanten Klassenbreite $2c$.

Unter diesen Annahmen gilt

$$\bar{x} - \bar{z} = 0$$

und

$$\hat{s}_x^2 - \hat{s}_z^2 = 4c^2 / 12 \quad 1)$$

In der Praxis findet man häufig, vor allem bei ökonomischen Untersuchungen, asymmetrische Verteilungen vor. Auch lassen sie sich nicht unbedingt auf ein bestimmtes Intervall beschränken.

Ich will deshalb durch ein Modell unter allgemeinen Verteilungsannahmen den Fehler, der durch die Klasseneinteilung bei Stichproben entsteht, untersuchen und quantifizieren.

1) Vgl.: Fersch, F.: Deskriptive Statistik, a.a.O., S. 97
Spiegel, M.R.: Theory and Problems of Statistik, New York 1972, S. 72, 81, 82, 90 und 95
und: Stange, K.: Angewandte Statistik, a.a.O., S. 78 ff.

3. Zusammenfassung

Der erste Teil der Arbeit zeigt, daß Klasseneinteilung nicht nur zur Vereinfachung von Stichproben dient, sondern sie kommt auch in vielerlei Hinsicht bei der Datenerfassung zur Anwendung. Es steht außer Frage: Die Gruppenbildung verursacht einen Informationsverlust.

Um eine statistische Aussage zu treffen, nimmt die Literatur an, daß alle Werte innerhalb einer jeden Klasse in der zugehörigen Klassenmitte liegen. Hierzu gibt sie Faustregeln an, die die Klassenbildung zum Zweck der Vereinfachung von Stichproben betreffen.

Erhebungen - insbesondere Stichproben - werden im statistischen Sinne zur Hypothesenüberprüfung, zur Prognose oder zur Schätzung unbekannter Parameter verwendet. Deshalb ist es Ziel dieser Arbeit, diese Einflüsse der Klassenbildung zu untersuchen, wobei ich mich auf die Parameterschätzung beschränken will.

II. EINFLÜSSE AUF DIE PARAMETERSCHÄTZUNG

Wie schon im ersten Teil gesagt, sollen nun die Einflüsse der Klasseneinteilung bei Stichproben auf die Parameterschätzung untersucht werden. Hierzu ist es notwendig, die statistische Schätztheorie allgemein zu betrachten.

1. Allgemeine Schätztheorie

Ziel des statistischen Schätzens ist es, auf Grund eines Stichprobenergebnisses auf unbekannte Eigenschaften der Grundgesamtheit zu schließen, wobei die Stichprobe Teil der Grundgesamtheit ist. Die statistischen Schätzmethoden gehören damit zu den Inferenzverfahren, zu denen auch die Tests gezählt werden. Man spricht deshalb von Inferenz und nicht von Induktion, weil die gezogenen Schlüsse auf Basis eines empirischen Ergebnisses mit berechenbarer hoher Wahrscheinlichkeit gelten. Mit anderen Worten: Schätzverfahren setzen Beobachtungen voraus, die mittels einer Stichprobe gewonnen werden. Die Inferenzaussage stellt dann eine Wahrscheinlichkeitsaussage dar, die auf der Häufigkeitsinterpretation der Wahrscheinlichkeit beruht, wobei die Basis das Stichprobenergebnis bildet.

Stützt sich diese Aussage allein auf eine Stichprobe, so spricht man von der klassischen Inferenz. Im Gegensatz zur Bayes-Inferenz zum Beispiel basiert die Aussage nicht nur auf einem empirischen Ergebnis sondern zusätzlich auf Informationen,

den sogenannten Vorinformationen in Form einer a priori-Verteilung. Während bei den klassischen Schätzverfahren feste Parameterwerte vorausgesetzt werden, geht die Bayes-Schätzung von Parameterwerten aus, die vom Zufall beeinflusst werden und eine sogenannte a priori-Verteilung besitzen. Zusammen mit den Informationen aus der Stichprobe will man so zu besseren Schätzungen gelangen.¹⁾

Es ist also festzuhalten: Eine statistische Schätzung vollzieht sich auf der Basis einer Stichprobe unter Verwendung der Wahrscheinlichkeitsrechnung. Deshalb muß die Stichprobe bestimmte Eigenschaften erfüllen. Sie werden im folgenden kurz beschrieben.

1.1. Die Stichprobe

Gegeben sei eine Gesamtheit - die sogenannte Grundgesamtheit. Über sie will man Aufschluß bekommen. Es sind zum Beispiel bestimmte Parameter zu schätzen. Das Naheliegendste ist, alle Elemente der Grundgesamtheit zu untersuchen. In diesem Fall spricht man von einer Totalerhebung. Eine vollständige Erhebung und Aufbereitung einer großen statistischen Masse ist allerdings sehr kostspielig. Außerdem verstreicht zu viel Zeit bis die Ergebnisse vorliegen, und bei gedanklich unendlich großen Grundgesamtheiten ist eine Totalerhebung technisch

1) Vgl.: Rinne, H.: Statistik Teil 3, München 1979, S. 111. Zu Bayes-Verfahren vgl. auch.: Stange, K.: Bayes-Verfahren, Schätz- und Testverfahren bei Berücksichtigung von Vorinformationen, Berlin-Heidelberg-New York 1977.

nicht durchführbar. Deshalb erhebt man in der Praxis die Grundgesamtheit nur teilweise und spricht von Teilerhebung. Eine Teilerhebung soll möglichst genau die Grundgesamtheit widerspiegeln. Sie soll also repräsentativ sein. Man will ja von ihr auf die wesentlichen Eigenschaften der Grundgesamtheit schließen. Solche repräsentativen Teilgesamtheiten nennt man Stichproben.¹⁾

Sowohl die statistische Theorie als auch die Praxis haben im Laufe der Zeit eine Vielzahl von Verfahren des Ziehens von Stichproben entwickelt.²⁾ Im Rahmen dieser Arbeit soll nur ein kleiner Einblick gegeben werden, der keinen Anspruch auf Vollständigkeit erheben kann.

1.1.1. Einstufige Auswahl - Die uneingeschränkte Zufallsauswahl

Sind Erhebungs- und Auswahlseinheit identisch, so spricht man von einer einstufigen Stichprobe. Das Ziehen der Stichprobenwerte setzt direkt bei der Erhebungseinheit ein. Dieses setzt voraus, daß eine entsprechende Auswahlgrundlage in Form eines Verzeichnisses aller Erhebungseinheiten vorhanden

1) Vgl.: Statistisches Bundesamt (Hersg.): Stichproben in der amtlichen Statistik, Stuttgart-Mainz 1960, S. 13.

2) Zu Stichprobenverfahren vgl. u.a. auch: Chochran, W.G.: Stichprobenverfahren, Berlin 1972; Kellerer, H.: Theorie und Technik des Stichprobenverfahrens, in: Einzelschriften der Deutsch Statistischen Gesellschaft, Nr. 5, München 1974; Statistisches Bundesamt (Hersg.): Stichproben in der amtlichen Statistik, a.a.O.

ist. Nur so wird ein unmittelbarer Zugriff auf die einzelnen Erhebungseinheiten gestattet.

Äquivalent ist das Urnenmodell, durch das sich die uneingeschränkte Zufallsauswahl beschreiben läßt. Alle Ausprägungen der Grundgesamtheit befinden sich in einer Urne. Dann vollzieht sich die uneingeschränkte Zufallsauswahl entweder durch Ziehen mit Zurücklegen oder ohne Zurücklegen.

Beim Ziehen mit Zurücklegen geht man wie folgt vor: Der Inhalt der Urne wird gut gemischt. Dann zieht man das erste Element und legt es wieder in die Urne zurück. Nun wird der Inhalt erneut gemischt und das zweite Element gezogen. Nach dem Zurücklegen verfährt man mit den weiteren Zügen genauso. Durch diese Erhebungsform werden folgende stochastischen Eigenschaften sichergestellt:

- a) Die einzelnen Züge sind paarweise stochastisch unabhängig.
- b) Jedes Element der Grundgesamtheit hat die gleiche Wahrscheinlichkeit, in die Stichprobe zu gelangen.

Das Ziehen ohne Zurücklegen vollzieht sich im wesentlichen genauso, nur daß die gezogenen Elemente nicht wieder in die Urne zurückgelegt werden. Hierdurch sind die einzelnen Züge paarweise stochastisch abhängig.

Da sich das Ziehen mit Zurücklegen mathematisch einfacher behandeln läßt, geht man meist von diesem

Modell aus. Ist die Grundgesamtheit sehr groß, im Extremfall unendlich, oder der Stichprobenumfang im Verhältnis zur Grundgesamtheit sehr klein, so kann mathematisch das Ziehen ohne Zurücklegen wie das Ziehen mit Zurücklegen behandelt werden.

1.1.2. Mehrstufige Auswahl

Mehrstufige Stichproben zielen darauf ab, daß eine Hierarchie der Gesamtheit vorliegt. Eine zweistufige Stichprobe zum Beispiel vollzieht sich wie folgt: Die Grundgesamtheit wird in natürlich vorgegebene Teilgesamtheiten aufgeteilt. Aus diesen Teilgesamtheiten wird eine Zufallsstichprobe entnommen. Man erhält in der ersten Stufe eine Auswahl aller Teilgesamtheiten. Daran schließt sich die zweite Stufe an: Aus jeder ausgewählten Teilgesamtheit wird eine Zufallsstichprobe gezogen. Beispiele für zweistufige Stichproben sind: Die geschichtete Auswahl - Totalerhebung auf der ersten Stufe, Teilerhebung auf der zweiten Stufe - und die Klumpenauswahl - Teilerhebung auf der ersten Stufe, Totalerhebung auf der zweiten Stufe.

Es zeigt sich: Bei mehrstufigen Verfahren erhält man mehrere Stichproben, die alle durch uneingeschränkte Zufallsauswahl entstanden sind. Hier gehen in das Stichprobenverfahren vorhandene Informationen über die Grundgesamtheit mit ein. Man will somit zu besseren Ergebnissen kommen (vgl. Schichten- und Klumpeneffekt).

1.1.3. Die Stichprobe als mehrdimensionaler Zufallsvektor

Für eine Inferenzaussage - wie die Schätzung eines unbekannten Parameters - muß die Stichprobe bestimmte Eigenschaften erfüllen. Im folgenden soll davon ausgegangen werden, daß man die Stichprobe durch uneingeschränkte Zufallsauswahl aus einer sehr großen Grundgesamtheit entnimmt. Für die mathematische Analyse spielt es deshalb keine Rolle, ob sie durch Ziehen mit Zurücklegen oder durch Ziehen ohne Zurücklegen entsteht.

Es gilt demnach:

- a) Jedes Element der Grundgesamtheit hat die gleiche Wahrscheinlichkeit in Stichprobe zu gelangen.
- b) Die einzelnen Züge sind paarweise stochastisch unabhängig.

Gegeben ist nun eine Grundgesamtheit, deren Merkmal $Y : \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable darstellt und die Verteilungs-Funktion $F(y)$ besitzt. Aus dieser Grundgesamtheit soll eine Stichprobe vom Umfang n durch uneingeschränkte Zufallsauswahl gezogen werden.

Da jedes Element der Grundgesamtheit die gleiche Wahrscheinlichkeit besitzt, in die Stichprobe zu gelangen, ist die Realisation des ersten

Zuges Y_1 vor dem Ziehen eine Zufallsvariable mit den gleichen stochastischen Eigenschaften - der gleichen Verteilungsfunktion $F(y)$ - wie Y . Für die weiteren der n Züge gilt dies auch, denn sie sind alle paarweise stochastisch unabhängig. Damit besteht die Stichprobe vor dem Ziehen aus n Zufallsvariablen, die alle die gleiche Verteilung wie Y besitzen. Die Stichprobe kann man also als n -dimensionalen Zufallsvektor auffassen

$$(Y_1, Y_2, \dots, Y_n) \quad .$$

Seine Komponenten Y_i ($i = 1, 2, \dots, n$) besitzen die Verteilung $F(Y)$ der Zufallsvariablen Y .

Liegt eine konkrete Stichprobe nach dem Ziehen vor, so spricht man von einem Stichprobenvektor. Seine Komponenten werden mit kleinen Buchstaben gekennzeichnet

$$(y_1, y_2, \dots, y_n) \quad .$$

Dieser Stichprobenvektor stellt also eine realisierte Stichprobe dar mit den Elementen y_i ($i = 1, 2, \dots, n$).

Es läßt sich festhalten: Eine Stichprobe vom Umfang n ist vor dem Ziehen ein Zufallsvektor und wird durch große Buchstaben gekennzeichnet, nach dem Ziehen, wenn sie realisiert ist, spricht man von einem Stichprobenvektor und bezeichnet seine Elemente mit kleinen Buchstaben.

1.1.4. Die Stichprobenfunktion und ihre Eigenschaft

Wie schon gesagt, läßt sich die Stichprobe vor dem Ziehen als n -dimensionaler Zufallsvektor auffassen. Dieser soll für eine Inferenzaussage, zum Beispiel für das Schätzen eines unbekannten Parameters der Grundgesamtheit genutzt werden. Hierzu ist es notwendig, diesen Vektor in eine geeignete Form zu bringen. Man will damit erreichen, daß sich die ursprüngliche Dimension verringert. Es wird eine Stichprobenfunktion - auch Statistik genannt

$$G : \mathbb{R}^n \rightarrow \mathbb{R}$$

$$G = G(Y_1, Y_2, \dots, Y_n)$$

gebildet.

G stellt eine Rechenvorschrift für die Komponenten $Y_1, Y_2, \dots, Y_n$ dar. Als Funktion von n Zufallsvariablen - so setzt man voraus - ist G wieder eine Zufallsvariable.

Die Stichprobenfunktion hat im wesentlichen zwei Aufgaben zu erfüllen:

- 1) Sie soll zu einer Inferenzaussage genutzt werden.¹⁾

1) Vgl.: Rinne, H., Statistik Teil 3, a.a.O., S. 58.

- 2) Sie soll den n -dimensionalen Zufallsvektor in seiner Dimension verringern.

Gegeben ist eine Grundgesamtheit mit der Zufallsvariablen $Y : \Omega \rightarrow \mathbb{R}$. Die Verteilungsfunktion $F(y)$ von Y hängt von dem unbekannten Parameter θ ab. Er soll durch die Stichprobe vom Umfang n ($Y_1, Y_2, \dots, Y_n$) geschätzt werden. Die Stichprobenfunktion $G = G(Y_1, Y_2, \dots, Y_n)$ soll die Basis für die Schätzung des unbekannten Parameters θ bilden. Deshalb verlangt man von einer solchen Stichprobenfunktion die Eigenschaft der Suffizienz.

Definition : Eine Stichprobenfunktion

$G = G(Y_1, Y_2, \dots, Y_n)$ heißt suffizient für einen Parameter θ einer gegebenen Verteilung $F(y)$, wenn die bedingte Verteilung des n -dimensionalen Zufallsvektors $(Y_1, Y_2, \dots, Y_n)$ bei gegebenem Wert von g nicht mehr von dem Parameter θ abhängt. Ist dies der Fall, dann hat die Stichprobe $(Y_1, Y_2, \dots, Y_n)$ keine weitere Informationen über θ .¹⁾

Mit Hilfe des Faktorisierungskriteriums²⁾ läßt sich die Suffizienz für eine Stichprobenfunktion G bei gegebener Verteilungsfunktion $F(y)$ für den unbekannten Parameter θ nachweisen.

1) Vgl.: Bickel, P.J. u. Doksum, K.A.: Mathematical Statistics, San Francisco-Düsseldorf-Johannesburg 1977, S. 63.

2) Vgl.: Ebenda, S. 65.

Zum Beispiel: Entsteht eine Stichprobe $(Y_1, Y_2, \dots, Y_n)$ aus einer exponentialverteilten Grundgesamtheit mit dem unbekannten Parameter λ , dann ist

$$G = \sum_{i=1}^n Y_i$$

suffizient für λ .

Oder: Entsteht die Stichprobe $(Y_1, Y_2, \dots, Y_n)$ aus einer normalverteilten Grundgesamtheit mit den unbekannten Parametern μ und σ^2 , dann ist

$$G_1 = \sum_{i=1}^n Y_i$$

suffizient für μ und

$$G_2 = \sum_{i=1}^n Y_i^2$$

suffizient für σ^2 .¹⁾

1) Vgl.: Bickel, P.J. u. Doksum, K.A.: Mathematical Statistics, San Francisco-Düsseldorf-Johannesburg 1977, S. 63 u. 66.

1.2. Definition der Schätzfunktion

Eine Schätzfunktion hat die Aufgabe, unbekannte Parameter einer Verteilung der Grundgesamtheit zu schätzen.

Die Zufallsvariable Y der Grundgesamtheit habe die Verteilungsfunktion $F(y)$ von Y . Die Lage und die Form der Verteilungsfunktion ist abhängig von den unbekannten Parametern

$\theta_1, \theta_2, \dots, \theta_1, \dots, \theta_m$. Sie sollen durch eine Stichprobe vom Umfang n ($Y_1, Y_2, \dots, Y_n$), die man durch uneingeschränkte Zufallsauswahl gewinnt, geschätzt werden. Hierzu konstruiert man Schätzfunktionen (Schätzer) der Form

$$\hat{\theta}_l = \hat{\theta}_l(Y_1, Y_2, \dots, Y_n)$$

$$l = 1, 2, \dots, m.$$

Die Schätzfunktion stellt also eine Rechenvorschrift mit den Stichprobenelementen dar. Vor dem Ziehen ist die Stichprobe bekanntlich ein n -dimensionaler Zufallsvektor. Als Funktion von n Zufallsvariablen soll die Schätzfunktion $\hat{\theta}_l$ wiederum eine Zufallsvariable mit einer Verteilungsfunktion $F(\hat{\theta}_l)$ darstellen.

Berechnet man die Schätzfunktion auf Basis der realisierten Stichprobe ($y_1, y_2, \dots, y_n$), so ergibt sich aus

$$\hat{\theta}_l = \hat{\theta}_l(y_1, y_2, \dots, y_n)$$

ein ganz bestimmter reeller Wert. Er wird als Schätzwert für den unbekannten Parameter θ_1 benutzt. In diesem Fall spricht man von Punktschätzung, weil dem Parameter auf Grund eines konkreten Stichprobenergebnisses ein bestimmter Wert zugewiesen wird.

Damit der Fehler bei einer Schätzung eines unbekannten Parameters möglichst gering wird, verlangt man von einer Schätzfunktion bestimmte stochastische Eigenschaften. Sie will ich im folgenden betrachten.

1.2.1. Erwartungstreue Schätzfunktion

Eine Schätzfunktion $\hat{\theta}_1 = \hat{\theta}_1(Y_1, Y_2, \dots, Y_n)$ heißt erwartungstreu für θ_1 , wenn gilt

$$E(\hat{\theta}_1) = \theta_1.$$

Die Erwartungstreue ist damit eine Eigenschaft der endlichen Stichproben. Sie besagt, daß man durch das Mittel aller von einer erwartungstreuen Schätzfunktion gelieferten Schätzwerte den wahren Parameter θ_1 erhält.

Beispiel:

Die Zufallsvariable Y habe die beliebige Verteilungsfunktion $F(y)$. Der Erwartungswert $E(Y) = \mu_y$ und die Varianz $\text{Var}(Y) = \sigma_y^2$ mögen existieren. Diese beiden Parameter seien unbekannt. Das arithmetische Mittel

$$\bar{Y} = 1/n \sum_{i=1}^n Y_i$$

der Stichprobenelemente $Y_1, Y_2, \dots, Y_n$ ist erwartungstreu für μ_y . Es gilt

$$E(\bar{Y}) = E(Y) = \mu_y .$$

Und die Stichprobenvarianz

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2$$

ist erwartungstreu für σ_y^2 .

Auch hier ist wieder

$$E(S^2) = \text{Var}(Y) = \sigma_y^2 .$$

Gilt für eine Schätzfunktion $\hat{\theta}_1$, wenn der Stichprobenumfang n über alle Schranken wächst,

$$\lim_{n \rightarrow \infty} E(\hat{\theta}_1) = \theta_1 ,$$

dann heit $\hat{\theta}_1$ asymptotisch erwartungstreu fr θ_1 .

1.2.2. Konsistente Schtzfunktion

Von einer "guten" Schtzfunktion $\hat{\theta}_1$ erwartet man, da bei Erhhung des Stichprobenumfangs n der Schtzwert besser wird, also der Schtzwert $\hat{\theta}_1$ nher an dem wahren Parameter θ_1 liegt.

Eine Schtzfunktion $\hat{\theta}_1$ heit konsistent fr θ_1 , wenn fr ein beliebig kleines $\epsilon > 0$ gilt

$$\lim_{n \rightarrow \infty} P (|\hat{\theta}_1 - \theta_1| < \epsilon) = 1 .$$

Um die Konsistenzeigenschaft nachzuweisen, mu man die Verteilung von θ_1 kennen, deshalb greift man fr den Nachweis oft auf folgenden Satz zurck.

Satz: Wenn $\hat{\theta}_1$ erwartungstreu fr θ_1 ist, und fr die Varianz von θ_1 gilt

$$\lim_{n \rightarrow \infty} \text{Var}(\hat{\theta}_1) = 0 ,$$

dann erfllt $\hat{\theta}_1$ auch die Eigenschaft der Konsistenz.

Beispiel:

Folgt die Zufallsvariable Y der beliebigen Verteilungsfunktion $F(y)$ mit den Parametern

$E(Y) = \mu_y$ und $\text{Var}(Y) = \sigma_y^2$, dann sind - wie aus dem letzten Abschnitt bekannt - $\bar{Y}$ und S^2 erwartungstreue Schätzfunktionen für μ_y bzw. σ_y^2 .

Die Varianz von $\bar{Y}$ ergibt sich nun aus

$$\text{Var}(\bar{Y}) = \frac{\text{Var}(Y)}{n} .$$

Da $\lim_{n \rightarrow \infty} \text{Var}(\bar{Y}) = 0$

besitzt $\bar{Y}$ auch die Eigenschaft der Konsistenz.

Das Gleiche ergibt sich auch für die Stichprobenvarianz S^2 .

1.2.3. Effiziente Schätzfunktion

Ein von einer Schätzfunktion $\hat{\theta}_1$ gelieferter Schätzwert θ_1 stimmt selten mit dem wahren Parameter θ_1 überein. Wenn es sich bei $\hat{\theta}_1$ um eine erwartungstreue Schätzfunktion handelt, dann hängt die Treffsicherheit dieser Schätzfunktion von der Varianz von $\hat{\theta}_1$ ab.

Erfüllen nun $\hat{\theta}_{11}$ und $\hat{\theta}_{21}$ die Eigenschaft der Erwartungstreue für θ_1 und gilt

$$\text{Var}(\hat{\theta}_{11}) < \text{Var}(\hat{\theta}_{21}),$$

dann heißt $\hat{\theta}_{11}$ effizienter als $\hat{\theta}_{21}$.

Diese Definition der Effizienz verdeutlicht, daß zwei erwartungstreue Schätzfunktionen mit ihrer Treffsicherheit verglichen werden.

Der Wunsch eines jeden Statistikers ist es, bei der Schätzung von unbekannten Parametern möglichst Schätzfunktionen zu finden, die die kleinste Varianz unter allen erwartungstreuen Schätzfunktionen besitzen. Erfüllt nun $\hat{\theta}_1$ diese Eigenschaft, dann wird sie absolut effizient für θ_1 genannt.

Beispiel:

Entstammt eine Stichprobe ($Y_1, Y_2, \dots, Y_n$) aus einer normalverteilten Grundgesamtheit mit $E(Y) = \mu_y$ und $\text{Var}(Y) = \sigma_y^2$, so läßt sich mit Hilfe der Ungleichung von Rao-Cramer nachweisen :

$$\bar{Y} = 1/n \sum_{i=1}^n Y_i$$

ist erwartungstreu für θ_1 und gleichzeitig absolut effiziente Schätzfunktion.

Und falls μ bekannt ist, gelten für

$$\hat{S}^2 = 1/n \sum_{i=1}^n (Y_i - \mu)^2$$

die Eigenschaften der Erwartungstreue und der absoluten Effizienz.¹⁾

-
- 1) Vgl. hierzu und zu den anderen Eigenschaften der Schätzfunktionen u.a.:
- Bamberg, G. u. Baur, F.: Statistik, München-Wien 1980, S. 147 ff.
 - Bleymüller, G. u.a.: Statistik für Wirtschaftswissenschaftler, a.a.O., S. 85 ff.
 - Bethea, R.M. u.a.: Statistical Methods for Engineers and Scientists, New York-Basel 1975, S. 119 ff.
 - Chou, Ya-lun: Statistical Analysis, London-New York-Sydney-Toronto 1969, S. 273 ff.
 - Fisz, M.: Wahrscheinlichkeitsrechnung und mathematische Statistik, Berlin 1970, S. 536.
 - Freund, J.E.: Mathematical Statistics, London 1972, S. 254.
 - Gänssler, P. u. Stute, W.: Wahrscheinlichkeitstheorie, Berlin-Heidelberg-New York 1977, S. 239 ff..
 - Hays, W.L. u. Winkler R.L.: Statistics: Probability, Inference and Decision, New York-Chicago-San Francisco ... 1971, S. 307 ff.
 - Rinne, H.: Statistik Teil 3, a.a.O., S. 88 ff.
 - Sachs, L.: Angewandte Statistik, Berlin-Heidelberg-New York 1978, S. 55 ff.
 - Wonnacott, T.H. u. Wonnacott, R.J.: Introductory Statistics for Business and Economics, New York-London-Sydney-Toronto 1972, S. 141 ff.
 - Zacks, S.: The Theory of Statistical Inference, New York-London-Sydney-Toronto 1971, S. 100 ff.

1.3. Zusammenfassung

Im letzten Abschnitt wurde gezeigt, wie eine statistische Schätzung abläuft: Die Zufallsvariable Y folgt einer beliebigen Verteilungsfunktion $F(y)$ von Y . Ihr Verlauf und ihre Form hängt von den unbekannten Parametern

θ_1 ($1 = 1, 2, \dots, m$) ab. Sie sollen auf Basis einer Stichprobe geschätzt werden.

Hierzu wird aus dieser Grundgesamtheit eine Stichprobe vom Umfang n durch uneingeschränkte Zufallsauswahl entnommen. Die Stichprobenelemente sind vor dem Ziehen alle Zufallsvariablen, die die gleichen stochastischen Eigenschaften besitzen wie die Zufallsvariable Y der Grundgesamtheit. Mit diesen n Stichprobenelementen bildet man eine Schätzfunktion $\hat{\theta}_1$. Sie stellt als Funktion von n Zufallsvariablen wieder eine zufällige Veränderliche dar. Auf Basis der realisierten Stichprobe erhält man dann den Schätzwert $\hat{\theta}_1$ für den wahren Parameter θ_1 .

Erfüllt nun die Schätzfunktion $\hat{\theta}_1$ die Eigenschaften der Erwartungstreue, der Konsistenz und der absoluten Effizienz, dann ist der Fehler der Schätzung bei gegebenem Stichprobenumfang minimal.

Um die Einflüsse der Klasseneinteilung auf die Parameterschätzung zu analysieren, muß untersucht werden, wie sich die Gruppenbildung auf die

entsprechenden Schätzfunktionen auswirkt.
Das heißt: Entstehen durch die Klassenbildung neue Schätzfunktionen, die sogar andere Parameter abschätzen? Wenn dies der Fall ist, kann durch Vergleich der beiden Schätzfunktionen aus den gruppierten Daten und den Schätzfunktionen aus den Originaldaten - der Informationsverlust untersucht und analysiert werden.

Hierzu ist es notwendig, die Klasseneinteilung durch ein mathematisches Modell zu beschreiben.

2. Das Modell der Klasseneinteilung

2.1. Modellannahmen

Gegeben sei eine Zufallsvariable Z mit der beliebigen stetigen Verteilungsfunktion $F(z)$. Die Dichte $f(z) = F'(z)$ soll existieren. Die Lage und Form der Verteilungsfunktion möge von den unbekannten Parametern θ_{z1} ($1 = 1, 2, \dots, m$) abhängen. Sie sollen durch eine Stichprobe vom Umfang n geschätzt werden. Eine uneingeschränkte Zufallsauswahl liefert den n -dimensionalen Zufallsvektor $(Z_1, Z_2, \dots, Z_n)$. Seine Komponenten Z_i besitzen alle die gleichen stochastischen Eigenschaften wie die Veränderliche Z , denn sie sind paarweise unabhängig.

Seien nun $\hat{\theta}_{z1} = \hat{\theta}_{z1}(Z_1, Z_2, \dots, Z_n)$ suffiziente Stichprobenfunktionen für die Parameter θ_{z1} . Als Schätzfunktionen erfüllen sie die Eigenschaften der Erwartungstreue, der Konsistenz und der absoluten Effizienz.¹⁾ Das heißt, auf der Basis einer Stichprobe mit gegebenem Stichprobenumfang n sind die Fehler der Schätzungen minimal.

Aus den früher genannten Gründen können die Realisationen z_i ($i = 1, 2, \dots, n$) von Z nicht beobachtet werden. Es ist eine Klasseneinteilung vorzunehmen. Hierzu unterteilt man den gesamten Definitionsbereich der Dichte $f(z)$ in abzählbar unendlich viele paarweise punktfremde Teilintervalle der Form

$$[z'_j; z'_{j+1}) \quad (j = \dots, -2, -1, 0, 1, 2, \dots).$$

1) Vgl.: Teil II, Abschnitte 1.1.4. u. 1.2..

Es werden gleiche Klassenbreiten unterstellt.
Das heißt, es gilt

$$2c_j = 2c = \text{const}$$

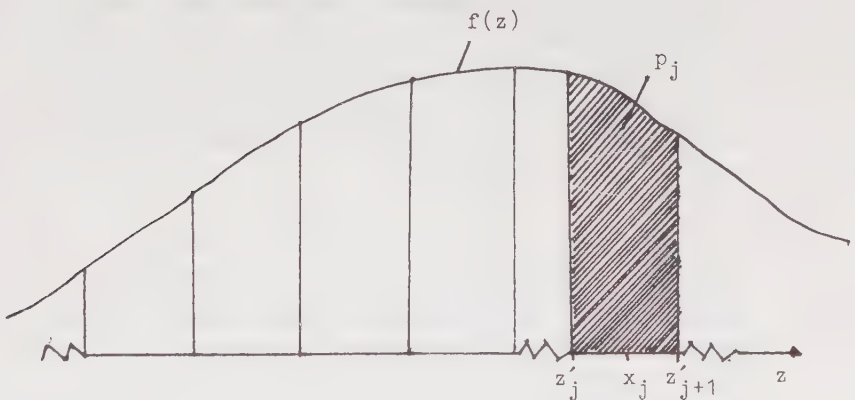
$$\text{für } j = \dots, -2, -1, 0, 1, 2, \dots$$

Diese Voraussetzung dient nur der Rechenvereinfachung und kann in der späteren Analyse leicht wieder aufgehoben werden.

Tritt nun in der Stichprobe die nicht beobachtbare Realisation z_i auf, für die die Ungleichung

$$z'_j \leq z_i < z'_{j+1}$$

erfüllt ist, so wird durch die Klassenbildung die Klassenmitte x_j beobachtet.



Wie die obige Abbildung zeigt, geschieht dies mit

der Wahrscheinlichkeit

$$p_j = \int_{z_j}^{z_{j+1}'} f(z) dz \quad .$$

Die x_j ($j = \dots, -2, -1, 0, 1, 2, \dots$) stellen damit alle möglichen Realisationen einer neuen Zufallsvariablen X dar, die durch die Klasseneinteilung aus der stetig verteilten Veränderlichen Z entstanden ist.

Fazit:

Die Klassenbildung transformiert eine stetig verteilte Veränderliche Z in eine diskret verteilte Zufallsvariable X .

2.2. Die diskrete Verteilung der Variablen X und ihre Parameter

Die aus der Variablen Z entstandene Veränderliche X hat folgende Eigenschaften:

Die Zufallsvariable X besitzt die Realisationen x_j ($j = \dots, -2, -1, 0, 1, 2, \dots$). Die Wahrscheinlichkeit, daß $X = x_j$ ist, hängt von der Dichte $f(z)$ sowie der Lage des Intervalls

$$[z_j', z_{j+1}')$$

ab und hat den oben definierten Wert p_j .

Somit besitzt X die Wahrscheinlichkeitsfunktion

$$\psi(x) = \begin{cases} p_j & \text{für } X = x_j \\ 0 & \text{sonst} \end{cases}$$

$j = \dots, -2, -1, 0, 1, 2, \dots$

$F(x)$ sei dann die dazugehörige Verteilungsfunktion von X . Ihre Lage und Form möge von den Parametern θ_{x1} ($1 = 1, 2, \dots, m$) abhängen, wobei θ_{x1} die gleiche Eigenschaft von $F(x)$ angibt wie θ_{z1} von $F(z)$. Zum Beispiel sei μ_x der Erwartungswert $E(X)$ von X , dann ist μ_z der Erwartungswert $E(Z)$ von Z .

Durch die Klassenbildung beobachtet man in der Stichprobe nun nicht die Realisationen z_i von Z , sondern Realisationen der Veränderlichen X . Das heißt, durch uneingeschränkte Zufallsauswahl erhält man den n -dimensionalen Zufallsvektor $(X_1, X_2, \dots, X_n)$. Seine Komponenten haben dann die gleichen stochastischen Eigenschaften wie die Zufallsvariable X .

Es seien nun

$$\hat{\theta}_{x1} = \hat{\theta}_{x1}(X_1, X_2, \dots, X_n)$$

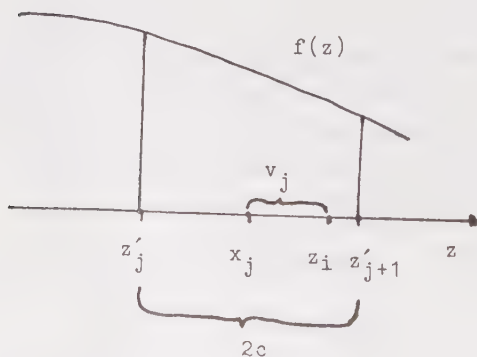
suffiziente Stichprobenfunktionen für die Parameter θ_{x1} , und als Schätzfunktionen mögen sie die Eigenschaften der Erwartungstreue, der Konsistenz und der absoluten Effizienz für die unbekannten Parameter θ_{x1} erfüllen. Damit werden durch klassifizierte

Daten bei gegebenem Stichprobenumfang die Parameter θ_{x1} der Verteilungsfunktion $F(x)$ der Zufallsvariablen X mit minimalem Fehler angenähert. Gesucht sind aber die Parameter θ_{z1} der Verteilungsfunktion $F(z)$ der Veränderlichen Z . Die Parameter θ_{z1} werden bekanntlich nur durch die Originaldaten z_i mit minimalem Fehler geschätzt.

Nun läßt sich der Informationsverlust, der durch die Klassenbildung entsteht, durch Vergleich der Parameter θ_{x1} von X und der Parameter θ_{z1} von Z analysieren. Hierzu ist es notwendig, den Zusammenhang zwischen der Zufallsvariablen Z und der Veränderlichen X genauer zu beleuchten.

2.3. Der Zusammenhang zwischen den Zufallsvariablen X und Z

Man betrachte eine beliebige Klasse j mit den Klassengrenzen z'_j und z'_{j+1} sowie der Klassenmitte x_j .



Innerhalb dieser Klasse j möge die Zufallsvariable Z den Wert z_j realisieren. Nach der Definition der Klassenbildung beobachtet man dann die Klassenmitte x_j . Wie die obige Abbildung zeigt, gilt die Beziehung

$$z_j = x_j + v_j \quad (-c \leq v_j < c) .$$

v_j stellt damit den Unterschied zwischen der nicht beobachtbaren Größe z_j und der beobachtbaren Größe x_j dar. Innerhalb der Klasse j können nun verschiedene z -Werte auftreten, deren Anzahl die absolute Klassenhäufigkeit bildet. Dadurch erscheinen auch unterschiedliche v_j -Werte. Da die Verteilung $F(z)$ von Z als stetig vorausgesetzt wird, lassen sich die v_j als Realisationen einer stetig verteilten Zufallsvariablen V_j auffassen. Somit kann die Veränderliche Z durch

$$(1) \quad Z = x_j + V_j$$

für alle x_j ($j = \dots, -2, -1, 0, 1, 2, \dots$).

ausgedrückt werden.

Durch die Anwendung dieser Beziehung (1) kann die Differenz zwischen den Parametern θ_{x1} von X und den Parametern θ_{z1} von Z bestimmt werden und damit der Informationsverlust, der durch die Klasseneinteilung entsteht. Dies soll im folgenden für den Erwartungswert $E(Z)$ und für die Varianz $\text{Var}(Z)$ einer stetig verteilten Zufallsvariablen Z gezeigt werden.

3. Der Einfluß der Klasseneinteilung auf den Erwartungswert $E(Z)$ und die Varianz $\text{Var}(Z)$ einer beliebigen stetigen Verteilung $F(z)$

3.1. Die stetige Verteilung $F(z)$ und ihre Parameter $E(Z)$ und $\text{Var}(Z)$

Gegeben sei die Zufallsvariable Z mit der stetigen Verteilungsfunktion $F(z)$ und der Dichte $f(z)$.

Der Erwartungswert $E(Z) = \mu_Z$ und die Varianz $\text{Var}(Z) = \sigma_Z^2$ mögen existieren. Damit gelten folgende Beziehungen:

$$f(z) \geq 0 \quad \text{für alle } z \text{ aus } \mathbb{R}$$

$$\int_{-\infty}^{\infty} f(z) \, dz = 1$$

$$F(z) = \int_{-\infty}^z f(t) \, dt$$

$$E(Z) = \int_{-\infty}^{\infty} z f(z) \, dz = \mu_Z$$

$$\begin{aligned} \text{Var}(Z) &= \int_{-\infty}^{\infty} (z - \mu_Z)^2 f(z) \, dz = \\ &= \int_{-\infty}^{\infty} z^2 f(z) \, dz - \mu_Z^2 = \sigma_Z^2 \end{aligned}$$

Die Parameter $E(Z)$ und $\text{Var}(Z)$ seien unbekannt. Sie sollen durch eine Stichprobe vom Umfang n geschätzt

werden. Die Stichprobe $(Z_1, Z_2, \dots, Z_n)$ stellt bekanntlich vor dem Ziehen einen n -dimensionalen Zufallsvektor dar, und die Schätzfunktionen

$$\bar{Z} = 1/n \cdot \sum_{i=1}^n Z_i \quad \text{für } E(Z)$$

und

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (Z_i - \bar{Z})^2 \quad \text{für } \text{Var}(Z)$$

erfüllen die Eigenschaften der Erwartungstreue und Konsistenz für $E(Z)$ bzw. $\text{Var}(Z)$.¹⁾

Da aus den genannten Gründen die Realisationen z_i nicht beobachtet werden können, führt man eine Klasseneinteilung durch. Die reelle z -Achse wird in abzählbar unendlich viele paarweise punktfremde Intervalle $[z'_j, z'_{j+1})$ mit den konstanten Breiten $2c$ unterteilt.

Durch die Klassenbildung entsteht - wie gesagt - aus der stetig verteilten Zufallsvariable Z eine diskret verteilte Veränderliche X .

3.2. Die diskrete Verteilung $F(x)$ und ihre Parameter $E(X)$ und $\text{Var}(X)$

Durch die Klasseneinteilung ist aus der stetig verteilten Zufallsvariable Z eine diskret verteilte

1) Vgl.: Teil II, Abschnitte 1.2.1 und 1.2.2.

Veränderliche X geworden. Sie besitzt die Wahrscheinlichkeitsfunktion

$$\psi(x) = \begin{cases} p_j & \text{für } X = x_j \quad (j = \dots, -2, -1, 0, 1, 2, \dots) \\ 0 & \text{sonst} \end{cases}.$$

Der Erwartungswert $E(X)$ von X

$$E(X) = \sum_{j=-\infty}^{\infty} x_j p_j = \mu_x$$

und die Varianz $\text{Var}(X)$ von X

$$\text{Var}(X) = \sum_{j=-\infty}^{\infty} (x_j - \mu_x)^2 p_j = \sum_{j=-\infty}^{\infty} x_j^2 p_j - \mu_x^2 = \sigma_x^2$$

existieren.

Durch die Klasseneinteilung erhält man nun eine Stichprobe vom Umfang n mit den beobachteten Größen $(X_1, X_2, \dots, X_n)$. Sie stellt vor dem Ziehen einen n -dimensionalen Zufallsvektor dar, dessen Komponenten die gleichen stochastischen Eigenschaften wie X besitzen. Aus II, 1. dieser Arbeit ist bekannt, daß

$$\bar{X} = 1/n \sum_{i=1}^n X_i$$

erwartungstreu und konsistent für μ_x ist sowie die Varianz

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

die Erwartungstreue und Konsistenz für σ_x^2 erfüllt.

In einer konkreten Stichprobe sind eventuell mehrere der Ausprägungen x_1 gleich x_j (mögliche Realisationen von X). Es treten also in der Stichprobe die x_j -Werte mit den absoluten Häufigkeiten f_j auf, wobei nur endlich viele eine Häufigkeit von $f_j > 0$ besitzen. Die realisierte Stichprobe liefert dann die Schätzwerte

$$\bar{x} = 1/n \sum_{j=-\infty}^{\infty} x_j f_j \quad \text{für } \mu_x$$

und

$$s^2 = \frac{1}{n-1} \sum_{j=-\infty}^{\infty} (x_j - \bar{x})^2 f_j \quad \text{für } \sigma_x^2.$$

Durch die klassifizierte Stichprobe werden also die Parameter $E(X)$ und $\text{Var}(X)$ der diskret verteilten Veränderlichen X angenähert. Eigentlich sind aber die Parameter $E(Z)$ und $\text{Var}(Z)$ der stetig verteilten Zufallsvariablen Z zu schätzen. Sie aber können nur durch die nicht beobachtbaren Realisationen z_i von Z angenähert werden.

Der Informationsverlust, der durch die Klasseneinteilung entsteht, kann damit durch Vergleich des Erwartungswertes $E(Z)$ und des Erwartungswertes $E(X)$ sowie durch Vergleich der Varianz $\text{Var}(Z)$ und der Varianz $\text{Var}(X)$ analysiert werden.

Gilt nämlich

$$E(Z) = E(X)$$

und

$$\text{Var}(Z) = \text{Var}(X),$$

so hat die Klasseneinteilung keinen Informationsverlust zur Folge.

Da die diskrete Verteilung $F(x)$ durch die Gruppenbildung aus der stetigen Verteilung $F(z)$ entstanden ist, kann man über die Beziehung (1)

$$Z = x_j + V_j$$

für alle x_j ($j = \dots, -2, -1, 0, 1, 2, \dots$)

den Unterschied zwischen den Parametern untersuchen. Hierzu ist es notwendig, einige Sätze über bedingte Erwartungswerte darzustellen.

3.3. Bedingte Verteilungen, bedingte Erwartungswerte und bedingte Varianzen

3.3.1. Die bedingte Wahrscheinlichkeitsalgebra und die bedingte Verteilungsfunktion

Definition: Bedingte Wahrscheinlichkeitsalgebra

Gegeben sei die Menge Ω mit den Elementarereignissen ω . A' sei eine σ -Ereignisalgebra und gleichzeitig Teilmenge der Potenzmenge von Ω . Ihre Elemente werden mit A, B , usw. bezeichnet. Ferner sei B' ein nicht leeres Mengensystem mit $B' \subset A'$. Es existiert nun die Mengenfunktion $P(A/B)$ für $A \in A'$ und $B \in B'$. Sie wird bedingte Wahrscheinlichkeit des Ereignisses A unter der Bedingung B genannt.

Es mögen nun folgende Axiome gelten:

$$1) P(A/B) \geq 0$$

$$P(B/B) = 1$$

$$\text{für } A \in A' \text{ und } B \in B'.$$

- 2) Ist $B \in B'$ fest, so ist $P(A/B)$ als Mengenfunktion von A auf A' ein Maß, das heißt für $A_k \in A'$
 $k = 1, 2, \dots$ und $A_k \cap A_{k+1} = \emptyset$. gilt

$$P\left(\sum_{k=1}^{\infty} (A_k/B)\right) = \sum_{k=1}^{\infty} P(A_k/B) \quad .$$

3) Es sei $A \in \mathcal{A}'$, $B \in \mathcal{B}'$ und $B \subseteq C$

$$P(A/B) = \frac{P(AB/C)}{P(B/C)} \quad 1)$$

Wenn diese Axiome Gültigkeit besitzen, heißt das System $(\Omega, \mathcal{A}', \mathcal{B}', P(A/B))$ bedingte Wahrscheinlichkeitsalgebra mit

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

und

$$P(B) > 0 \quad 2)$$

Definition: Bedingte Verteilungsfunktion und bedingte Dichte

Sei nun $Y: \Omega \rightarrow \mathbb{R}$ eine beliebige Zufallsvariable und B ein Ereignis, das mit positiver Wahrscheinlichkeit eintritt und zur Wahrscheinlichkeitsalgebra von Y gehört. Unter der bedingten Verteilungsfunktion von Y bezüglich der Bedingung B versteht man die Funktion

$$F(y/B) := P(Y \leq y/B) \quad .$$

Ist $F(y/B)$ absolut stetig differenzierbar, so heißt

$$f(y/B) = F'(y/B)$$

1) $AB = A \cap B$

2) Vgl.: Rényi, A.: Wahrscheinlichkeitsrechnung
Berlin 1962, S. 56.

bedingte Dichte von Y bezüglich der Bedingung B.

Falls $B=\Omega$ ist, gilt

$$f(y/B) = f(y)$$

und

$$F(y/B) = F(y) .$$

Stellt B_k ($k = 1, 2, \dots$) ein vollständiges System von Ω dar, so erhält man die Verteilungsfunktion von Y durch

$$F(y) = \sum_{k=1}^{\infty} P(B_k) F(y/B_k)$$

(Satz der totalen Wahrscheinlichkeit)

und die Dichte von Y durch

$$f(y) = \sum_{k=1}^{\infty} P(B_k) f(y/B_k) . \quad 1)$$

Sei nun $Y: \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable mit der stetigen Verteilungsfunktion $F(y)$. y'_k bezeichnet das Ereignis, daß Y in das Intervall $[y_k; y_{k+1})$ fällt. Die Wahrscheinlichkeit, daß Y gleich $Y = y'_k$ ist, ergibt sich aus

$$\begin{aligned} P(y'_k) &= P(y_k \leq Y < y_{k+1}) = F(y_{k+1}) - F(y_k) = \\ &= \int_{y_k}^{y_{k+1}} f(y) dy , \end{aligned}$$

1) Vgl.: Rényi, A.: Wahrscheinlichkeitsrechnung, a.a.O., S. 151.

wobei $f(y)$ die Dichte von Y darstellt.

Mit $h(y, y'_k)$ bezeichnet man die gemeinsame Dichte des Zufallsvektors (Y, Y'_k) . Die bedingte Dichte von Y bezüglich der Bedingung y'_k ist dann gleich

$$f(y/y'_k) = \frac{h(y, y'_k)}{P(y'_k)}$$

für $P(y'_k) > 0$.

Die bedingte Verteilungsfunktion von Y bezüglich der Bedingung $Y = y'_k$ berechnet sich aus

$$F(y/y'_k) = \int_{y_k}^y f(t/y'_k) dt. \quad 1)$$

Definition: Bedingte Erwartung

Sei $Y: \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable und $F(y)$ ihre Verteilungsfunktion bzw. $f(y)$ ihre Dichte auf $\mathbb{R}$. Das Intervall $[y_k, y_{k+1})$ sei Teil von $\mathbb{R}$ und wird von Y mit positiver Wahrscheinlichkeit angenommen. Dann ist der bedingte Erwartungswert von Y bezüglich der Bedingung $y_k \leq Y \leq y_{k+1}$ gleich

$$E(Y/y_k \leq Y \leq y_{k+1}) = \frac{\int_{y_k}^{y_{k+1}} y f(y) dy}{\int_{y_k}^{y_{k+1}} f(y) dy} \quad 1)$$

1) Vgl. Rényi, A.: Wahrscheinlichkeitsrechnung, a.a.O., S. 211.

3.3.2. Eigenschaften der bedingten Erwartung

Man betrachte die bedingte Erwartung von Y bezüglich der Bedingung $y_k \leq Y \leq y_{k+1}$

$$E(Y/y_k \leq Y \leq y_{k+1})$$

und bezeichnet $Y = y'_k$ für $y_k \leq Y \leq y_{k+1}$, so läßt sich die obige Aussage auch schreiben als

$$E(Y/y'_k) .$$

Es sei nun y'_k für $k = 1, 2, \dots$ ein vollständiges System von $Y: \Omega \rightarrow \mathbb{R}$. Damit läßt sich die bedingte Erwartung von Y bezüglich der Bedingung y'_k $E(Y/y'_k)$ als Realisation einer neuen Zufallsgröße

$$E(Y/Y')$$

auffassen, wobei ihre Realisationen $E(Y/Y' = y'_k)$ mit den Wahrscheinlichkeiten $P(y'_k)$ auftreten.

Es gilt nun für den Erwartungswert der bedingten Erwartung

$$E[E(Y/Y')] = E(Y) ,$$

falls $E(Y)$ existiert. ¹⁾

1) Vgl.: Vogel, W.: Wahrscheinlichkeitstheorie; Göttingen 1970, S. 141/142.

Es sei nun $E(Y)$ der Erwartungswert der Zufallsvariablen Y und $\text{Var}(Y)$ die Varianz von Y . Die Varianz von $E(Y/Y')$ ist gleich

$$\text{Var}[E(Y/Y')] = E[(E(Y/Y'))^2] - [E(E(Y/Y'))]^2.$$

Die Varianz von Y selbst läßt sich ausdrücken durch

$$\text{Var}(Y) = \text{Var}[E(Y/Y')] + E[\text{Var}(Y/Y')],$$

wobei $\text{Var}(Y/Y' = y'_k)$ die bedingte Varianz von Y bezüglich der Bedingung $Y' = y'_k$ darstellt. Sie berechnet sich aus

$$\begin{aligned}\text{Var}(Y/Y' = y'_k) &= E[(Y - E(Y/Y' = y'_k))^2 / Y' = y'_k] \\ &= E(Y^2/Y' = y'_k) - [E(Y/Y' = y'_k)]^2. \quad 1)\end{aligned}$$

1) Vgl.: Rényi, A.: Wahrscheinlichkeitsrechnung, a.a.O., S. 233.

3.4. Der Einfluß der Klasseneinteilung auf den Erwartungswert $E(Z)$

Es sei $Z: \Omega \rightarrow \mathbb{R}$ wieder eine Zufallsvariable mit der stetigen Dichte $f(z)$. Der Erwartungswert $E(Z) = \mu_Z$ und die Varianz $\text{Var}(Z) = \sigma_Z^2$ von Z sollen existieren. Die Klassenbildung unterteilt die gesamte reelle z -Achse in abzählbar unendlich viele Teilintervalle $[z'_j; z'_{j+1})$ mit den konstanten Breiten

$$2c = z'_{j+1} - z'_j \quad j = \dots, -2, -1, 0, 1, 2, \dots$$

Die Klassenmitten ergeben sich aus

$$x_j = \frac{z'_{j+1} + z'_j}{2} \quad j = \dots, -2, -1, 0, 1, 2, \dots$$

Somit ist die untere Intervallgrenze gleich

$$z'_j = x_j - c$$

und die obere Intervallgrenze gleich

$$z'_{j+1} = x_j + c$$

Der Erwartungswert

$$\begin{aligned} E(Z) &= \int_{-\infty}^{\infty} z f(z) dz \\ &= \mu_Z \quad \text{von } Z \end{aligned}$$

kann damit auch ausgedrückt werden durch

$$(2) \quad E(Z) = \sum_{j=-\infty}^{\infty} \int_{x_j - c}^{x_j + c} z f(z) dz .$$

Wie aus II, 2.1. bekannt ist, entsteht durch die Klassenbildung aus der stetig verteilten Variablen Z eine diskret verteilte Veränderliche X mit den Realisationen x_j ($j = \dots, -2, -1, 0, 1, 2, \dots$). Zwischen diesen beiden Zufallsvariablen besteht wie gesagt die Beziehung (1)

$$Z = x_j + V_j$$

$$\text{für alle } x_j \text{ (} j = \dots, -2, -1, 0, 1, 2, \dots \text{). } ^1)$$

Da im Ausdruck (2) nur über die Klasse j integriert wird, kann z durch $z = x_j + v_j$ substituiert werden;

$$\text{aus } z = x_j + v_j$$

$$\text{folgt : } dz = dv_j ,$$

für die untere Grenze

$$z = x_j - c$$

$$\text{folgt: } v_j = -c$$

1) Vgl.: Teil II, Abschnitt 2.3..

und für die obere Grenze

$$z = x_j + c$$

$$\text{folgt: } v_j = + c \quad ;$$

damit erhält man für den Erwartungswert

$$E(Z) = \sum_{j=-\infty}^{\infty} \int_{-c}^{+c} (x_j + v_j) f(x_j + v_j) dv_j \quad .$$

Da x_j innerhalb der Klasse j eine feste Größe ist, kann $E(Z)$ umgeschrieben werden in

$$E(Z) = \sum_{j=-\infty}^{\infty} x_j \int_{-c}^{+c} f(x_j + v_j) dv_j +$$

$$\sum_{j=-\infty}^{\infty} \int_{-c}^{+c} v_j f(x_j + v_j) dv_j .$$

Im ersten Summanden des obigen Ausdrucks läßt sich $x_j + v_j$ durch z substituieren. Das heißt, aus

$$x_j + v_j = z$$

ergibt sich

$$dv_j = dz \quad ,$$

für die untere Grenze

$$v_j = -c$$

$$\text{folgt: } z = x_j - c$$

und für die obere Grenze

$$v_j = +c$$

$$\text{folgt: } z = x_j + c.$$

Somit erhält man

$$E(Z) = \sum_{j=-\infty}^{\infty} x_j \int_{x_j - c}^{x_j + c} f(z) dz +$$

$$\sum_{j=-\infty}^{\infty} \int_{-c}^{+c} v_j f(x_j + v_j) dv_j.$$

Da

$$\int_{x_j - c}^{x_j + c} f(z) dz = p_j$$

die Wahrscheinlichkeit ist, daß Z in die Klasse $[z'_j; z'_{j+1})$ fällt, ergibt

$$E(Z) = \sum_{j=-\infty}^{\infty} x_j p_j + \sum_{j=-\infty}^{\infty} \int_{-c}^{+c} v_j f(x_j + v_j) dv_j.$$

Nun ist

$$\sum_{j=-\infty}^{\infty} x_j p_j = E(X)$$

gleich dem Erwartungswert $E(X)$ der diskret verteilten Veränderlichen X .

Damit erhält man für den Erwartungswert $E(Z)$ von Z

$$(3) \quad E(Z) = E(X) + \sum_{j=-\infty}^{\infty} \int_{-c}^{+c} v_j f(x_j + v_j) dv_j.$$

Es läßt sich nun festhalten:

Durch die Klasseneinteilung verändert sich der Erwartungswert $E(Z)$ von Z um die Größe

$$\sum_{j=-\infty}^{\infty} \int_{-c}^{+c} v_j f(x_j + v_j) dv_j$$

zum Erwartungswert $E(X)$ von X .

Im folgenden soll der Ausdruck

$$\sum_{j=-\infty}^{\infty} \int_{-c}^{+c} v_j f(x_j + v_j) dv_j$$

genauer untersucht werden. Hierzu betrachtet man

$$\int_{-c}^{+c} v_j f(x_j + v_j) dv_j .$$

Es läßt sich wieder $x_j + v_j$ durch z substituieren;

$$\text{aus } x_j + v_j = z$$

$$\text{folgt: } dv_j = dz ,$$

für die untere Grenze

$$v_j = -c$$

$$\text{folgt: } z = x_j - c$$

und für die obere Grenze

$$v_j = c$$

$$\text{folgt: } z = x_j + c .$$

Somit ergibt sich

$$\int_{-c}^{+c} v_j f(x_j + v_j) dv_j =$$

$$\int_{x_j-c}^{x_j+c} (z - x_j) f(z) dz =$$

$$\int_{x_j-c}^{x_j+c} z f(z) dz - x_j \int_{x_j-c}^{x_j+c} f(z) dz .$$

Dividiert man nun obige Gleichung durch

$$\int_{x_j-c}^{x_j+c} f(z) dz ,$$

so erhält man

$$\frac{\int_{-c}^{+c} v_j f(x_j + v_j) dv_j}{\int_{x_j-c}^{x_j+c} f(z) dz} =$$

$$\frac{\int_{x_j - c}^{x_j + c} z f(z) dz}{\int_{x_j - c}^{x_j + c} f(z) dz} = x_j .$$

Die Klassen $[z'_j; z'_{j+1})$ stellen ein vollständiges System der Zufallsvariablen $Z: \Omega \rightarrow \mathbb{R}$ dar. Da sie von Z mit positiver Wahrscheinlichkeit angenommen werden, ist

$$\frac{\int_{x_j - c}^{x_j + c} z f(z) dz}{\int_{x_j - c}^{x_j + c} f(z) dz} = E(Z/x_j)$$

gleich dem bedingten Erwartungswert von Z bezüglich der Bedingung $X = x_j$.

Damit ergibt sich

$$\frac{\int_{-c}^{+c} v_j f(x_j + v_j) dv_j}{\int_{x_j - c}^{x_j + c} f(z) dz} = E(Z/x_j) - x_j .$$

$E(Z/x_j) - x_j$ läßt sich auffassen als die erwartete Fehlergröße von V_j in der Klasse j - also als bedingte Erwartung von V bezüglich der Bedingung x_j

$$E(Z/x_j) - x_j = E(V/x_j).$$

Das gesuchte Integral ist gleich

$$(4) \quad \int_{-c}^{+c} v_j f(x_j + v_j) dv_j = E(V/x_j) p_j.$$

Durch Einsetzen von (4) in (3) ergibt sich

$$(5) \quad E(Z) = E(X) + \sum_{j=-\infty}^{+\infty} E(V/x_j) p_j.$$

x_j ($j = \dots, -2, -1, 0, 1, 2, \dots$) stellt ein vollständiges System von Z dar. Hieraus folgt, daß

$$\sum_{j=-\infty}^{\infty} E(V/x_j) p_j = E[E(V/X)]$$

gleich dem Erwartungswert der bedingten Erwartung von V bezüglich der Bedingung X ist.¹⁾

1) Vgl.: Teil II, Abschnitt 3.3.2.

Hieraus ergibt sich für (5) :

$$(6) \quad E(Z) = E(X) + E[E(V/X)] = \mu_Z.$$

Fazit:

Durch die Klasseneinteilung verändert sich der Erwartungswert $E(Z)$ von Z um die Größe $E[E(V/X)]$ zum Erwartungswert $E(X)$ von X .

Dieses Ergebnis behält auch seine Gültigkeit, wenn unterschiedliche Klassenbreiten $2c_j$ ($j = \dots, -2, -1, 0, 1, 2, \dots$) unterstellt werden. Die bedingte Erwartung von V bezüglich der Bedingung $X = x_j$ berücksichtigt nach ihrer Definition die jeweilige Klassenbreite der Klasse j . Die Voraussetzung konstanter Klassenbreiten kann somit aufgegeben werden.

3.5. Der Einfluß der Klasseneinteilung auf die Varianz $\text{Var}(Z)$

Die Zufallsvariable Z habe wieder die gleichen Eigenschaften wie in II 3.4.. Auch die Klasseneinteilung vollzieht sich nach dem bekannten Schema. Die Varianz $\text{Var}(Z)$ von Z ergibt sich aus

$$\text{Var}(Z) = \sigma_Z^2 = \int_{-\infty}^{+\infty} z^2 f(z) dz - \mu_Z^2.$$

Durch die Klassenbildung läßt sie sich umschreiben in

$$(7) \quad \text{Var}(Z) = \sum_{j=-\infty}^{\infty} \int_{x_j-c}^{x_j+c} z^2 f(z) dz - \mu_z^2 .$$

Von Beziehung (7) betrachtet man zuerst den Ausdruck

$$\sum_{j=-\infty}^{\infty} \int_{x_j-c}^{x_j+c} z^2 f(z) dz .$$

Da auch hier wieder nur über die Klasse j integriert wird, kann z durch $x_j + v_j$ substituiert werden,

$$\text{aus } z = x_j + v_j$$

$$\text{folgt: } dz = dv_j ,$$

für die untere Grenze

$$z = x_j - c$$

$$\text{folgt: } v_j = -c$$

und für die obere Grenze

$$z = x_j + c$$

$$\text{folgt: } v_j = c .$$

Damit ergibt sich

$$(8) \quad \sum_{j=-\infty}^{\infty} \int_{x_j-c}^{x_j+c} z^2 f(z) dz =$$

$$\sum_{j=-\infty}^{\infty} \int_{-c}^{+c} (x_j + v_j)^2 f(x_j + v_j) dv_j =$$

$$\sum_{j=-\infty}^{\infty} \int_{-c}^{+c} x_j^2 f(x_j + v_j) dv_j +$$

$$\sum_{j=-\infty}^{\infty} \int_{-c}^{+c} 2x_j v_j f(x_j + v_j) dv_j +$$

$$\sum_{j=-\infty}^{\infty} \int_{-c}^{+c} v_j^2 f(x_j + v_j) dv_j =$$

$$\sum_{j=-\infty}^{\infty} x_j^2 \int_{-c}^{+c} f(x_j + v_j) dv_j +$$

$$\sum_{j=-\infty}^{\infty} 2x_j \int_{-c}^{+c} v_j f(x_j + v_j) dv_j +$$

$$\sum_{j=-\infty}^{\infty} \int_{-c}^{+c} v_j^2 f(x_j + v_j) dv_j .$$

Im ersten Summanden des obigen Ausdrucks läßt sich $x_j + v_j$ durch z substituieren; das heißt,

$$\text{aus } z = x_j + v_j$$

$$\text{folgt: } dz = dv_j ,$$

für die untere Grenze

$$v_j = -c$$

$$\text{folgt: } z = x_j - c$$

und für die obere Grenze

$$v_j = c$$

$$\text{folgt: } z = x_j + c.$$

Damit ergibt sich für (8)

$$(9) \quad \sum_{j=-\infty}^{\infty} \int_{x_j-c}^{x_j+c} z^2 f(z) dz =$$

$$\sum_{j=-\infty}^{\infty} x_j^2 \int_{x_j-c}^{x_j+c} f(z) dz +$$

$$\sum_{j=-\infty}^{\infty} 2x_j \int_{-c}^{+c} v_j f(x_j + v_j) dv_j +$$

$$\sum_{j=-\infty}^{\infty} \int_{-c}^{+c} v_j^2 f(x_j + v_j) dv_j =$$

$$\sum_{j=-\infty}^{\infty} x_j^2 p_j + \sum_{j=-\infty}^{\infty} 2x_j \int_{-c}^{+c} v_j f(x_j + v_j) dv_j +$$

$$\sum_{j=-\infty}^{\infty} \int_{-c}^{+c} v_j^2 f(x_j + v_j) dv_j.$$

Durch die Beziehung (4) (II 3.4.) erhält man dann aus (9)

$$(10) \quad \sum_{j=-\infty}^{\infty} \int_{x_j-c}^{x_j+c} z^2 f(z) dz =$$

$$\sum_{j=-\infty}^{\infty} x_j^2 p_j + 2 \sum_{j=-\infty}^{\infty} x_j E(V/x_j) p_j +$$

$$\sum_{j=-\infty}^{\infty} \int_{-c}^{+c} v_j^2 f(x_j + v_j) dv_j .$$

In (10) betrachtet man nun das Integral des letzten Summanden

$$\int_{-c}^{+c} v_j^2 f(x_j + v_j) dv_j .$$

Hier läßt sich wieder $x_j + v_j$ durch z substituieren;

$$\text{aus } z = x_j + v_j$$

$$\text{folgt: } dz = dv_j ,$$

für die untere Grenze

$$v_j = -c$$

$$\text{folgt: } z = x_j - c$$

und für die obere Grenze

$$v_j = c$$

$$\text{folgt: } z = x_j + c.$$

Damit ist

$$(11) \quad \int_{-c}^{+c} v_j^2 f(x_j + v_j) dv_j =$$

$$\int_{x_j-c}^{x_j+c} (z - x_j)^2 f(z) dz =$$

$$\int_{x_j-c}^{x_j+c} z^2 f(z) dz - 2x_j \int_{x_j-c}^{x_j+c} z f(z) dz +$$

$$x_j^2 \int_{x_j-c}^{x_j+c} f(z) dz \quad .$$

Durch Division durch

$$\int_{x_j-c}^{x_j+c} f(z) dz$$

erhält man

$$\frac{\int_{-c}^{+c} v_j^2 f(x_j + v_j) dv_j}{\int_{x_j-c}^{x_j+c} f(z) dz} =$$

$$\frac{\int_{x_j-c}^{x_j+c} z^2 f(z) dz}{\int_{x_j-c}^{x_j+c} f(z) dz} - 2x_j \frac{\int_{x_j-c}^{x_j+c} z f(z) dz}{\int_{x_j-c}^{x_j+c} f(z) dz} + x_j^2$$

(a) (b)

Audruck (b) ist wieder die bedingte Erwartung $E(Z/x_j)$ von Z bezüglich der Bedingung $X = x_j$. ¹⁾

Nun zum Ausdruck (a): Die Varianz von Z existiert. Die Zufallsvariable $Z: \Omega \rightarrow \mathbb{R}$ nimmt die Klasse $j [z'_j, z'_{j+1})$, die Teil von $\mathbb{R}$ ist, mit positiver Wahrscheinlichkeit an. Somit ist Ausdruck (a) gleich der bedingten Erwartung $E(Z^2/x_j)$ von Z^2 bezüglich der Bedingung $X = x_j$. ²⁾

Damit ergibt sich für

$$\frac{\int_{-c}^{+c} v_j^2 f(x_j + v_j) dv_j}{\int_{x_j-c}^{x_j+c} f(z) dz} = E(Z^2/x_j) - 2x_j E(Z/x_j) + x_j^2.$$

1) Vgl.: Teil II, Abschnitt 3.4..

2) Vgl.: Teil II, Abschnitt 3.2.2..

Mit Hilfe der Sätze über Erwartungswerte erhält man

$$(12) \quad \frac{\int_{-c}^{+c} v_j f(x_j + v_j) dv_j}{x_j + c} = \int_{x_j - c} f(z) dz$$

$$E[Z^2 - 2x_j Z + x_j^2/x_j] =$$

$$E[(Z - x_j)^2/x_j] =$$

$$E(V^2/x_j) \quad .$$

Mit (12) folgt für (11)

$$(13) \quad \int_{-c}^{+c} v_j f(x_j + v_j) dv_j = E(V^2/x_j)p_j \quad .$$

Mit (10), (7) und (13) bestimmt sich die Varianz von Z aus

$$(14) \quad \text{Var}(Z) = \sum_{j=-\infty}^{\infty} x_j^2 p_j +$$

$$2 \sum_{j=-\infty}^{\infty} x_j E(V/x_j) p_j + \quad ,$$

$$\sum_{j=-\infty}^{\infty} E(V^2/x_j) p_j - \mu_Z^2 \quad .$$

Es gilt

$$\sum_{j=-\infty}^{\infty} x_j^2 p_j = E(X^2) \quad .$$

Da x_j ($j = \dots, -2, -1, 0, 1, 2, \dots$) ein vollständiges System von $Z: \Omega \rightarrow \mathbb{R}$ bilden, ergibt

$$\sum_{j=-\infty}^{\infty} E(V^2/x_j) p_j = E[E(V^2/X)]$$

und

$$\sum_{j=-\infty}^{\infty} x_j E(V/x_j) p_j = E[X E(V/X)] \quad .$$

Damit wird aus (14)

$$(15) \quad \text{Var}(Z) = E(X^2) + 2E[X E(V/X)] +$$

$$E[E(V^2/X)] - \mu_z^2.$$

Ab diesem Ergebnis kann die Voraussetzung gleicher Klassenbreiten aufgegeben werden. Denn: Sowohl die bedingte Erwartung von V bezüglich der Bedingung $X = x_j$ als auch die bedingte Erwartung von V^2 bezüglich der Bedingung $X = x_j$ berücksichtigen nach ihren Definitionen die jeweilige Klassenbreite $2c_j$ der Klasse j .

In Gleichung (15) muß nun noch μ_z durch Gleichung (6) (II 3.4.) ersetzt werden, so ergibt die Beziehung (15)

$$(16) \quad \text{Var}(Z) = E(X^2) + 2E[X E(V/X)] +$$

$$E[E(V^2/X)] - [E(X) + E[E(V/X)]]^2.$$

Durch Auflösen der letzten Klammer erhält man

$$\begin{aligned} \text{Var}(Z) &= E(X^2) + 2E[X E(V/X)] + \\ &E[E(V^2/X)] - [E(X)]^2 - 2E(X)E[E(V/X)] - \\ &[E[E(V/X)]]^2 \end{aligned}$$

oder

$$(17) \quad \text{Var}(Z) = E(X^2) - [E(X)]^2 + \\ 2E[X E(V/X)] - 2E(X) E[E(V/X)] + \\ E[E(V^2/X)] - [E[E(V/X)]]^2 .$$

In Gleichung (17) ist

$$E(X^2) - [E(X)]^2 = \text{Var}(X)$$

gleich der Varianz von X .

Und da x_j ($j = \dots, -2, -1, 0, 1, 2, \dots$) ein vollständiges System von $Z: \Omega \rightarrow \mathbb{R}$ ist, gilt

$$E[E(V/X)] = E(V)$$

und

$$E[E(V^2/X)] = E(V^2) .$$

Mithin ist

$$E[E(V^2/X)] - [E[E(V/X)]]^2 = \text{Var}(V)$$

gleich der Varianz von V .

Nun soll noch der Ausdruck

$$E[X E(V/X)] - E(X) E[E(V/X)]$$

aus Gleichung (17) betrachtet werden.

$E(V/X)$ stellt bekanntlich eine Zufallsvariable mit den Realisationen $E(V/X = x_j)$ ($j = \dots, -2, -1, 0, 1, 2, \dots$) dar. X selbst ist ebenfalls eine zufällige Veränderliche mit den Realisationen $X = x_j$ ($j = \dots, -2, -1, 0, 1, 2, \dots$).

Hieraus folgt: Der obige Ausdruck ist gleich der Kovarianz des Zufallsvektors $(X; E(V/X))$.

Es gilt

$$\begin{aligned} E[X E(V/X)] - E(X) E[E(V/X)] &= \\ \text{Cov}[X; E(V/X)] &. \end{aligned}$$

Damit ergibt sich für (17)

$$(18) \quad \text{Var}(Z) = \text{Var}(X) + 2 \text{Cov}[X; E(V/X)] + \text{Var}(V) .$$

Fazit:

Durch die Klasseneinteilung verändert sich die Varianz $\text{Var}(Z)$ von Z um eine Größe

$$2\text{Cov}[X; E(V/X)] + \text{Var}(V)$$

zur Varianz $\text{Var}(X)$ von X , wobei Varianz von V durch die Beziehung

$$E[E(V^2/X)] - E[E(V/X)]^2 = \text{Var}(V)$$

berechnet werden kann.

Wie schon gesagt, behält dieses Ergebnis auch seine Gültigkeit, wenn unterschiedliche Klassenbreiten $2c_j$ ($j = \dots, -2, -1, 0, 1, 2, \dots$) unterstellt werden. Es ist nur notwendig, daß sich die Klasseneinteilung über den gesamten Definitionsbereich der Dichte $f(z)$ von Z vollzieht. Das heißt: Man kann die Randklassen offen lassen. Zum Beispiel die letzte Klasse läuft von einem Wert z'_k bis ∞

$[z'_k; \infty)$. Beispiele werden dies verdeutlichen. Als einzige Voraussetzung bleibt: Die Klassen $[z'_j; z'_{j+1})$ ($j = \dots, -2, -1, 0, 1, 2, \dots$) müssen ein vollständiges System von $Z: \Omega \rightarrow \mathbb{R}$ bilden.

3.6. Zusammenfassung der Ergebnisse

Gegeben sei wieder die Zufallsvariable $Z: \Omega \rightarrow \mathbb{R}$ mit der stetigen Dichte $f(z)$, dem Erwartungswert $E(Z)$ und der Varianz $\text{Var}(Z)$ von Z . Diese Parameter seien unbekannt und sollen durch eine Stichprobe vom Umfang n ($Z_1, Z_2, \dots, Z_n$) geschätzt werden. Aus den genannten Gründen können die Realisationen z_i ($i = 1, 2, \dots, n$) von Z nicht beobachtet werden. Man führt deshalb eine Klassenbildung auf der z -Achse durch. Die abgeleiteten Ergebnisse erlauben, wie gesagt, auch unterschiedliche Klassenbreiten. Die Klassen müssen nur ein vollständiges System von $\Omega \rightarrow \mathbb{R}$ bilden. Setzt man voraus, daß für die Dichte $f(z)$ gilt

$$f(z) > 0$$

für alle z aus $\mathbb{R}$,

das heißt, sie läuft von $-\infty$ nach $+\infty$ mit positiven Funktionswerten, so müssen die Klassen die gesamte reelle z -Achse überdecken. Da unterschiedliche Breiten erlaubt sind, können auch Klassen gebildet werden mit der Breite unendlich.

Wählt man nun $(k + t + 3)$ Klassen, so unterteilt man die reelle z -Achse von z'_{-k} bis z'_t in $(k + t + 1)$ punktfremde Intervalle der Form

$$[z'_j, z'_{j+1}) \quad (j = -k, \dots, -2, -1, 0, 1, 2, \dots, t).$$

Damit die Klassen ein vollständiges System $\Omega \rightarrow \mathbb{R}$ sind, bildet man noch eine rechte Randklasse

$$[z'_{t+1}, \infty)$$

sowie eine linke Randklasse

$$(-\infty, z'_{-k}) .$$

Sie werden im folgenden mit rechte bzw. linke offene Randklasse bezeichnet, während die Intervalle $[z'_j, z'_{j+1})$ geschlossene Klassen heißen.

Die Klassenmitten der geschlossenen Klassen ergeben sich dann wieder aus

$$x_j = \frac{z'_j + z'_{j+1}}{2}$$

$$j = -k, \dots, -2, -1, 0, 1, 2, \dots, t .$$

Die Klassenmitten x_{t+1} und x_{-k-1} können innerhalb der offenen Randklassen frei gewählt werden.

Durch diese Form der Gruppenbildung entsteht aus der stetig verteilten Zufallsvariablen Z eine

diskret verteilte zufällige Veränderliche X mit den Realisationen x_j ($k = -k, \dots, -2, -1, 0, 1, 2, \dots, t$), x_{t+1} und x_{-k-1} . Sie besitzt die Wahrscheinlichkeitsfunktion

$$\psi(x) = \begin{cases} p_{-k-1} & \text{für } X = x_{-k-1} \\ p_j & \text{für } X = x_j \quad (j = -k, \dots, t) \\ p_{t+1} & \text{für } X = x_{t+1} \\ 0 & \text{sonst} \end{cases},$$

wobei sich die jeweiligen Wahrscheinlichkeiten aus der Dichte $f(z)$ von Z ergeben.

Durch die Stichprobe der beobachtbaren Größen $(X_1, X_2, \dots, X_n)$ werden nun der Erwartungswert $E(X)$ und die Varianz $\text{Var}(X)$ der diskreten Zufallsvariablen X angenähert. Die gesuchten Parameter - der Erwartungswert $E(Z)$ und die Varianz $\text{Var}(Z)$ - der stetig verteilten zufälligen Veränderlichen Z können nur durch die nicht beobachtbaren Realisationen z_i geschätzt werden. Damit konnte der Fehler der Klasseneinteilung durch Vergleich der Parameter der stetig verteilten Zufallsvariablen Z mit denen der diskret verteilten zufälligen Veränderlichen X bestimmt werden.¹⁾

1) Vgl.: Teil II, Abschnitte 3.1. und 3.2..

Es ergab sich für den Erwartungswert

$$E(Z) = E(X) + E[E(V/X)] .$$

Damit wird $E(Z)$ bei Klasseneinteilung um den Betrag

$$E[E(V/X)]$$

- den Erwartungswert der bedingten Erwartung der Fehlervariablen V bezüglich der Bedingung X - zum Erwartungswert $E(X)$ der Zufallsvariablen X verändert.¹⁾ Diese oben genannte Größe heißt im folgenden Gesamtfehler des Erwartungswertes.

Und für die Varianz erhielt man

$$\text{Var}(Z) = \text{Var}(X) + 2\text{Cov}[X; E(V/X)] + \text{Var}(V) .$$

Damit verändert sich $\text{Var}(Z)$ durch die Klassenbildung um den Betrag

$$2\text{Cov} [X; E(V/X)] + \text{Var}(V)$$

- der Kovarianz der Zufallsvariablen X und $E(V/X)$ plus der Gesamtvarianz der Fehlervariablen V - zu der Varianz $\text{Var}(X)$ der diskreten Veränderlichen X .²⁾ Der oben genannte Ausdruck heißt im folgenden Gesamtfehler der Varianz.

1) Vgl.: Teil II, Abschnitt: 3.4..
Vgl.: Teil II, Abschnitt: 3.5..

Mit Hilfe dieser abgeleiteten Ergebnisse kann man den Einfluß einer gegebenen Klasseneinteilung bei bekannter Verteilungsform der Zufallsvariablen Z auf die Parameter $E(Z)$ und $\text{Var}(Z)$ quantifizieren und analysieren. Dies soll nun am Beispiel der Exponentialverteilung (Vertreterin schiefer Verteilungen) und der Laplace-Verteilung (Vertreterin symmetrischer Verteilungen) geschehen.

III. QUANTITATIVE ANALYSE

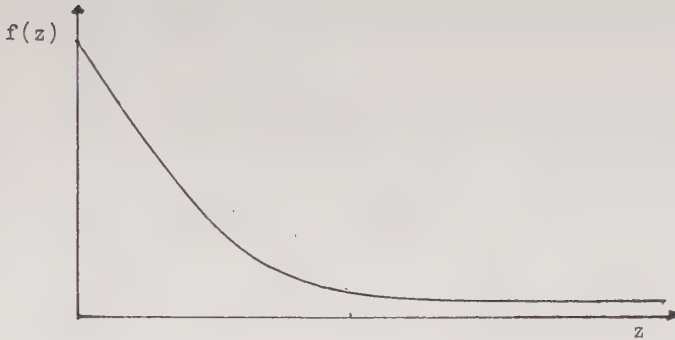
Es sollen nun die Gesamtfehlergrößen $E[E(V/X)]$ und $2\text{Cov}[X; E(V/X)] + \text{Var}(V)$ quantitativ bestimmt werden, um dann damit die Einflüsse der Klassenbildung zu interpretieren. Hierzu wird einmal unterstellt, daß die Zufallsvariable Z exponentialverteilt ist. Außerdem will ich die Einflüsse bei der Laplace-Verteilung untersuchen.

1. Eigenschaften der Exponentialverteilung

Die Exponentialverteilung wird häufig als Verteilung der Lebensdauer von technischen Produkten benutzt und kommt deshalb oft in der statistischen Qualitätskontrolle zur Anwendung. Sie besitzt die folgende Dichte

$$f(z) = \begin{cases} \lambda e^{-\lambda z} & \text{für } z \geq 0 \\ 0 & \text{sonst} \end{cases} .$$

Die Zufallsvariable Z folgt also der Exponentialverteilung mit der oben genannten Dichte $f(z)$



Der Erwartungswert $E(Z)$ ist dann gleich

$$E(Z) = 1/\lambda$$

und die Varianz $\text{Var}(Z)$ von Z ergibt sich aus

$$\text{Var}(Z) = 1/\lambda^2 .$$

Als weitere Eigenschaft der Exponentialverteilung sei hier genannt: Die Wahrscheinlichkeit, daß die Zufallsvariable Z einen Wert $Z \geq z_1$ annimmt, ist unabhängig von dem Ereignis: Z hat schon einen Wert $z_0 < z_1$ erreicht. Man sagt auch, die Exponentialverteilung ist "vergeßlich".

2. Die Eigenschaften der Laplace-Verteilung

Folgt die Zufallsvariable Z der Laplace-Verteilung, so hat sie die Dichte $f(z)$ von Z .

$$f(z) = \frac{1}{2\lambda} e^{-\frac{|x - \mu|}{\lambda}}$$

für alle z aus $\mathbb{R}$.

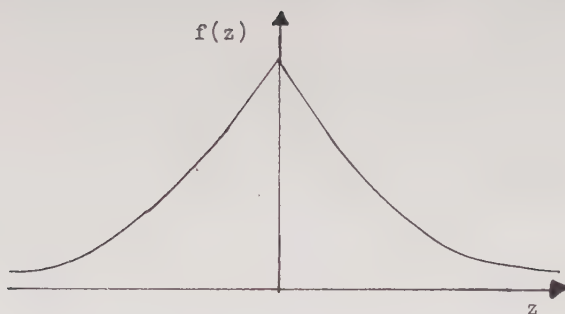
Der Erwartungswert $E(Z)$ ergibt sich in diesem Fall aus

$$E(Z) = \mu$$

und die Varianz $\text{Var}(Z)$ von Z durch

$$\text{Var}(Z) = 2\lambda^2$$

Die Dichte $f(z)$ der Laplace-Verteilung ist symmetrisch zu $z = \mu$ ihrem Erwartungswert und hat folgenden Graph



3. Klasseneinteilung bei exponentialverteilter Zufallsvariablen Z

Im Beispiel I (vgl. Anhang) wird die Exponentialverteilung mit $\lambda = 1$ bei folgender Klasseneinteilung untersucht:

- Es werden 13 Klassen gebildet.
- Das Intervall der z -Achse $[0;12)$ ist in 12 ($j = 0, 1, 2, \dots, 11$) Klassen mit den konstanten Breiten $2c = 1$ unterteilt.
- Die rechte Randklasse (Klasse 13) ist offen und läuft von 12 bis ∞ .

Die Klassenmitten der geschlossenen Klassen
ergeben sich aus

$$x_j = \frac{z'_j + z'_{j+1}}{2}$$

$$j = 0, 1, 2, \dots, 11.$$

Zum Beispiel ist

$$x_0 = 0,5.$$

Die Klassenmitte der offenen Randklasse
wurde mit

$$x_{13} = 12,5$$

festgelegt.

Die Berechnung der p_j erfolgt dann durch

$$p_j = \int_{z'_j}^{z'_{j+1}} e^{-z} dz ;$$

für die erste Klasse ergab sich damit ein

$$p_0 = 0,632121.$$

Für die Bestimmung der Fehlergröße des Erwartungswertes

$$E[E(V/X)] = \sum_{j=0}^{13} E(V/X=x_j) p_j$$

muß noch

$$E(V/X = x_j)$$

$$\text{für } j = 0, 1, 2, \dots, 13,$$

berechnet werden. Dies geschieht durch die Beziehung

$$E(V/X = x_j) = E(Z/X = x_j) - x_j$$

$$\text{für } j = 0, 1, 2, 3, \dots, 13 \quad .$$

Der bedingte Erwartungswert von Z bezüglich der Bedingung $X = x_j$ oder $z'_j \leq Z \leq z'_{j+1}$ $E(Z/X = x_j)$ läßt sich durch

$$E(Z/X=x_j) = \frac{\int_{z'_j}^{z'_{j+1}} ze^{-z} dz}{p_j} =$$

$$= \frac{e^{-z'_j}(z'_j + 1) - e^{-z'_{j+1}}(z'_{j+1} + 1)}{p_j}$$

$$\text{für } j=0, 1, 2, \dots, 13$$

berechnen.

Zum Beispiel erhält man für $x_0 = 0,5$ den Wert

$$E(Z/X = 0,5) = 0,41802$$

und mithin

$$E(V/X = 0,5) = -0,08198 .$$

Alle numerischen Werte sind im Anhang (Beispiel I) zusammengestellt. Für den Gesamtfehler des Erwartungswertes ergab sich

$$E[E(V/X)] = -0,081977 .$$

Das bedeutet: Der Erwartungswert $E(X)$ von X , der durch die klassifizierten Daten geschätzt wird, ist um 0,081977 größer als der gesuchte Erwartungswert $E(Z)$ von Z .

Nun ist der Gesamtfehler der Varianz

$$2\text{Cov}[X; E(V/X)] + \text{Var}(V)$$

zu berechnen.

Die Kovarianz kann leicht durch

$$\text{Cov} [X, E(V/X)] = \sum_{j=0}^{13} x_j E(V/X=x_j) p_j - E(X)E[E(V/X)]$$

bestimmt werden. Es ergab sich im Beispiel I (s. Anhang) ein Wert von

$$\text{Cov} [X; E(V/X)] = 0,000039 .$$

Die Berechnung der Varianz $\text{Var}(V)$ von V erfolgt über die Beziehung

$$\text{Var}(V) = E[\text{Var}(V/X)] + \text{Var}[E(V/X)] .^1)$$

Zur Bestimmung von $\text{Var}(V)$ überlegt man sich folgendes: $\text{Var}(V/X = x_j)$ stellt die bedingte Varianz der Fehlervariablen V bezüglich der Bedingung $X = x_j$ dar. Sie läßt sich auch schreiben als

$$\text{Var}(V/X = x_j) = \text{Var}(Z - x_j/X = x_j) .$$

Wenn die Bedingung $X = x_j$ erfüllt ist, ist x_j eine feste Größe. Damit gilt

$$\text{Var}(V/X = x_j) = \text{Var}(Z/X = x_j)$$

1) Vgl.: Teil II, Abschnitt 3.3.2..

und zugleich

$$E[\text{Var}(V/X)] = E[\text{Var}(Z/X)] .$$

Da sich die Varianz $\text{Var}(Z)$ von Z ausdrücken läßt durch

$$\text{Var}(Z) = E[\text{Var}(Z/X)] + \text{Var}[E(Z/X)] ,^{1)}$$

kann die Varianz $\text{Var}(V)$ von V leicht durch

$$\text{Var}(V) = \text{Var}(Z) - \text{Var}[E(Z/X)] + \text{Var}[E(V/X)]$$

berechnet werden.

Für das Beispiel I ergibt sich für die Varianz $\text{Var}(V)$ von V ein Wert von

$$\text{Var}(V) = 0,0793$$

und damit einen Gesamtfehler der Varianz von

$$2\text{Cov}[X; E(V/X)] + \text{Var}(V) = 0,079378 .$$

1) Vgl.: Teil II, Abschnitt 3.3.2..

Dies bedeutet: Die Varianz $\text{Var}(X)$ von X , die durch die klassifizierten Daten geschätzt wird, ist um 0,079378 kleiner als die gesuchte Varianz $\text{Var}(Z)$ von Z .

In den im Anhang dargestellten Beispielen II, III und IV wurde die Klassenzahl zuerst auf 8, dann auf 5 und schließlich auf 3 reduziert. Die Ergebnisse wurden durch die gleiche Vorgehensweise wie im Beispiel I gewonnen und sind dort ausgewiesen sowie in Tabelle (1) zusammengestellt.

4. Klasseneinteilung bei Laplace-verteilter Zufallsvariablen Z

Es soll nun die Laplace-Verteilung mit $E(Z) = \mu = 0$ und $\text{Var}(Z) = 32$ ($\lambda = 4$) untersucht werden. Damit besitzt die Zufallsvariable Z die Dichte $f(z)$ der Laplace-Verteilung

$$f(z) = \begin{cases} 1/8 e^{z/4} & \text{für } z \leq 0 \\ 1/8 e^{-z/4} & \text{für } z \geq 0 \end{cases} .$$

Die Dichte der Laplace-Verteilung ist bekanntlich auf der gesamten reellen z -Achse größer Null. Im Beispiel V wurde das Intervall $(-24; 24)$

der z-Achse in 12 Klassen ($j = -6, -5, \dots, -1, 0, 1, 2, \dots, 6$) mit den konstanten Breiten $2c = 4$ eingeteilt. Hinzu kommen noch eine linke offene Randklasse $(-\infty; -24)$ und eine rechte offene Randklasse $[24; \infty)$. Die Klassenmitten der offenen Randklassen wurden mit

$$x_{-7} = -26$$

und

$$x_7 = 26$$

festgelegt. Die Mitten der geschlossenen Klassen ergeben sich dann durch die bekannte Gleichung. Die Wahrscheinlichkeiten p_j ($j = -7, -6, \dots, 0, 1, 2, \dots, 6, 7$) können nach der gleichen Vorgehensweise wie die Beispiele in (III, 3) berechnet werden.

Zur Bestimmung des Gesamtfehlers des Erwartungswertes

$$E[E(V/X)]$$

benötigt man wieder den bedingten Erwartungswert von Z bezüglich der Bedingung $X = x_j$ bzw.

$$z'_j \leq Z \leq z'_{j+1} \quad E(Z/X = x_j).$$

Er ergibt sich aus unserer Dichte der Laplace-Verteilung durch

$$E(Z/X=x_j) = \begin{cases} \frac{\int_{z_j'}^{z_{j+1}'} 1/8 z e^{z/4} dz}{p_j} & \text{für } z \leq 0 \\ \frac{\int_{z_j'}^{z_{j+1}'} 1/8 z e^{-z/4} dz}{p_j} & \text{für } z \geq 0 \end{cases}$$

$$= \begin{cases} \frac{e^{z_{j+1}'/4} (1/2 z_{j+1}' - 2) - e^{z_j'/4} (1/2 z_j' - 2)}{p_j} & \text{für } j = -7, -6, \dots, -1 \\ \frac{e^{-z_j'/4} (1/2 z_j' - 2) - e^{-z_{j+1}'/4} (1/2 z_{j+1}' - 2)}{p_j} & \text{für } j = 0, 1, 2, \dots, 7 \end{cases}$$

Die numerischen Ergebnisse von $E(Z/X = x_j)$ und die hieraus folgenden von $E(V/X = x_j)$ sind im Anhang, Beispiel V, zusammengestellt.

Damit ist der Gesamtfehler des Erwartungswertes gleich

$$E[E(V/X)] = 0 \quad .$$

Dies bedeutet: Der Erwartungswert $E(X)$ von X , der durch die klassifizierten Daten geschätzt wird, ist gleich dem gesuchten Erwartungswert $E(Z)$ von Z

$$E(X) = E(Z) \quad .$$

Im Beispiel V hat die Klassenbildung also keinen Einfluß auf den Erwartungswert.

Der Gesamtfehler der Varianz

$$2\text{Cov}[X; E(V/X)] + \text{Var}(V)$$

kann hier analog der in III,3. beschriebenen Vorgehensweise berechnet werden. Das Beispiel V ergibt den Wert $-1,209739$.

Dies bedeutet: Die Varianz $\text{Var}(X)$ von X , die durch die klassifizierten Daten geschätzt wird, ist um $1,209739$ größer als die gesuchte Varianz $\text{Var}(Z)$ von Z .

Im Beispiel VI wurde die Anzahl der Klassen auf 8 reduziert. Danach (Beispiel VII) wurden 5 Klassen asymmetrisch angeordnet. Die negativen z -Werte werden durch eine Klasse $(-\infty; 0)$ repräsentiert, während die positive z -Achse in 4 Klassen aufgeteilt ist. Die einzelnen Ergebnisse sind im Anhang ausgewiesen sowie dort zusätzlich in Tabelle (2) zusammengestellt.

5. Interpretation der quantitativen Ergebnisse

5.1. Der Gesamtfehler des Erwartungswertes

5.1.1. Bei exponentialverteilter Zufallsvariablen Z

In Tabelle (1) (s. Anhang) wurde in der ersten Spalte für die verschiedenen Klasseneinteilungen der Gesamtfehler des Erwartungswertes $E[E(V/X)]$ zusammengestellt. Er ist immer negativ. Dies bedeutet: Der Erwartungswert $E(X)$ von X , der durch die klassifizierten Daten angenähert wird, ist immer größer als der gesuchte Erwartungswert $E(Z)$ von Z .

Man mag sich nun wundern, daß bei geringerer Klassenzahl der Gesamtfehler betragsmäßig kleiner wird. Um dies zu interpretieren, betrachtet man die Zahlen des Beispiels I (im Anhang). Hier weist $E(V/X = x_j)$ für alle x_j der geschlossenen Klassen einen konstanten negativen Wert auf.

(Dies liegt an der Eigenschaft der "Vergeßlichkeit" der Exponentialverteilung.) Nur der Wert der rechten offenen Randklasse $E(V/X = 12,5)$ ist positiv. Wenn man nun die Anzahl der Klassen verkleinert, geht dieser positive Wert mit höherer Wahrscheinlichkeit in den Gesamtfehler $E[E(V/X)]$ ein (Beispiele II - IV). Damit wird er betragsmäßig kleiner.

Hieraus kann man folgern: Der Gesamtfehler des Erwartungswertes ließe sich noch weiter verringern, wenn man die frei wählbare Klassenmitte an die linke Klassengrenze der offenen Randklasse legt.

5.1.2. Bei Laplace-verteilter Zufallsvariablen Z

In Tabelle (2) (s. Anhang) wurde in der ersten Spalte der Gesamtfehler $E[E(V/X)]$ bei Laplace-verteilter Zufallsvariablen Z zusammengestellt. Es zeigt sich: So lange die Klassen symmetrisch zu Null - dem Erwartungswert $E(Z) = 0$ der Laplace-Verteilung - angeordnet werden, hat die Klasseneinteilung keinen Einfluß auf den Erwartungswert. Das heißt, es gilt in diesen Fällen

$$E(X) = E(Z) .$$

Im Beispiel VII (asymmetrische Anordnung der Klassen) kommt es zu einem Gesamt-

fehler von

$$E[E(V/X)] = -1,106002 .$$

Bei dieser asymmetrischen Anordnung der Klassen ist der Erwartungswert $E(X)$ von X , der durch die klassifizierten Daten geschätzt wird, um 1,106002 größer als der gesuchte Erwartungswert $E(Z)$ von Z . Man kann sich nun leicht an Hand der Beispiele klarmachen, daß der Gesamtfehler bei asymmetrischer Anordnung auch positiv sein kann. Nämlich dann, wenn man auf der negativen z -Achse mehr Klassen wählt als auf der positiven. Für die betragsmäßige Größe gilt dann das gleiche, wie das, was bei exponentialverteiltem Z (III, 5.1.1.) gesagt wurde.

Es läßt sich also festhalten: Bei symmetrisch verteiltem Z ist der Gesamtfehler des Erwartungswertes Null, so lange man die Klassen symmetrisch zur Symmetrieachse anordnet. In allen anderen Fällen kann er positive oder negative Werte aufweisen.

5.2. Der Gesamtfehler der Varianz

5.2.1. Bei exponentialverteilter Zufallsvariablen Z

In Tabelle (1) (s. Anhang) wurde in der letzten Spalte der Gesamtfehler der Varianz bei exponentialverteilter Zufallsvariablen Z zusammengestellt. Man stellt fest, daß er immer größer Null ist und bei Verkleinerung der Klassenzahl steigt. Das heißt: Die Varianz $\text{Var}(X)$ von X , die durch die klassifizierten Daten geschätzt wird, ist hier immer kleiner als die gesuchte Varianz $\text{Var}(Z)$ von Z . Zur Interpretation des Phänomens betrachtet man sich die Zusammensetzung des Gesamtfehlers der Varianz:

$$2\text{Cov}[X; E(V/X)] + \text{Var}(V) .$$

Die Varianz $\text{Var}(V)$ der Fehlervariablen V berechnet sich bekanntlich aus

$$\text{Var}(V) = E[\text{Var}(V/X)] + \text{Var}[E(V/X)] .^1)$$

Im ersten Summanden läßt sich $\text{Var}(V/X)$ auffassen als die Varianz der Fehlervariablen V innerhalb der einzelnen Klassen, während sich $\text{Var}[E(V/X)]$ als Varianz von V außerhalb der Klassen interpretieren

1) Vgl.: Teil II, Abschnitt 3.2.2..

läßt. Wenn man diese Überlegung anstellt, leuchtet unmittelbar ein: Bei Verkleinerung der Klassenzahl steigt die Gesamtvarianz $\text{Var}(V)$ der Fehlervariablen V . Die Zahlen der Tabelle (1) dritte Spalte bestätigen das.

Weiter geht in den Gesamtfehler der Varianz die Kovarianz $\text{Cov}[X; E(V/X)]$ mit dem Faktor zwei ein. Die Kovarianz gibt den unnormierten Zusammenhang zwischen den Zufallsvariablen X und $E(V/X)$ an. Sie ist, wie Tabelle (1) zeigt, für alle Beispiele positiv und wird bei abnehmender Klassenzahl größer. Daß der unnormierte stochastische Zusammenhang größer Null ist, läßt sich durch die entsprechenden Werte des Beispiels I leicht erklären: Für alle geschlossenen Klassen ist $E(V/X = x_j) = -0,08198$ konstant. Nur für die offene Randklasse ist die bedingte Erwartung von V bezüglich der Bedingung $X = 12,5$ gleich $E(V/X = 12,5) = 0,5$ - also positiv.

Fazit:

Mit größer werdenden Werten von X nimmt $E(V/X)$ ebenfalls zu; damit ist der unnormierte stochastische Zusammenhang positiv und mithin

$$\text{Cov}[X; E(V/X)] > 0 .$$

Bei geringerer Klassenzahl (vgl. Beispiele II - IV) geht $E(V/X = x_j)$ schneller von negativen Beträgen in positive über. Die Kovarianz wird also größer.

Es kann somit festgehalten werden:

Da sowohl bei abnehmender Anzahl von Klassen die Kovarianz $\text{Cov}[X; E(V/X)]$ zunimmt als auch die Varianz $\text{Var}(V)$ der Fehlervariablen steigt, wird der Gesamtfehler der Varianz größer (vgl. Tabelle (1)).

Eine Bemerkung muß aber noch hinzugefügt werden: Bei der Interpretation des Gesamtfehlers des Erwartungswertes (III, 5.1.1.) wurde gesagt: Der Gesamtfehler des Erwartungswertes ließe sich verkleinern, wenn man die frei wählbare Klassenmitte der offenen Randklasse an die linke Klassengrenze legt. Dies hat aber Auswirkungen auf den Gesamtfehler der Varianz. Nämlich: Der Wert von $E(V/X=x_j)$ der Randklasse würde steigen, was eine Vergrößerung der Kovarianz $\text{Cov}[X; E(V/X)]$ nachsichzieht. Damit käme es zu einer Erhöhung des Gesamtfehlers der Varianz.

5.2.2. Bei Laplace-verteilter Zufallsvariablen Z

In Tabelle (2), letzte Spalte, ist der Gesamtfehler der Varianz bei Laplace-verteilter Zufallsvariablen Z zusammengestellt. Bei 14 Klassen weist er einen negativen Wert (-1,209739) auf. Das heißt, die Varianz $\text{Var}(X)$ von X, die durch die klassifizierten Daten geschätzt wird, ist in diesem Fall größer als die gesuchte Varianz $\text{Var}(Z)$ von Z. Im Beispiel VI (8 Klassen) dreht sich dann das Ergebnis um - die Varianz $\text{Var}(X)$ von X ist damit

kleiner als die Varianz $\text{Var}(Z)$ von Z . Wählt man schließlich eine asymmetrische Anordnung (z.B. 5 Klassen), wird der Gesamtfehler noch größer. Er kann damit - eine symmetrische Verteilung wie die Laplace-Verteilung unterstellt - sowohl positive als auch negative Werte aufweisen.

Um dies zu interpretieren, betrachtet man wieder die Zusammensetzung des Gesamtfehlers

$$2\text{Cov}[X; E(V/X)] + \text{Var}(V) .$$

Die Gründe für eine Erhöhung der Varianz $\text{Var}(V)$ von V bei geringer werdender Klassenzahl wurden in III, 5.2.1. ausführlich beschrieben. Da $\text{Var}(V)$ immer positiv ist, muß das oben beschriebene Phänomen an der Größe der Kovarianz liegen.

Die Kovarianz gibt bekanntlich den unnormierten Zusammenhang zwischen den Zufallsvariablen X und $E(V/X)$ an. Im Beispiel V treten die Werte $E(V/X = -26) = -2$ und $E(V/X = 26) = 2$ der Randklassen mit geringer Wahrscheinlichkeit auf. Sie können deshalb unberücksichtigt bleiben. Nun gilt (vgl. Beispiel V) : Bei Vergrößerung von X wird $E(V/X)$ kleiner. Damit wird der unnormierte Zusammenhang negativ.

Verringert man nun die Klassenzahl (vgl. Beispiel VI), gewinnen die Werte $E(V/X = x_j)$ der Randklassen an Bedeutung und die Kovarianz wird positiv.

Gleichzeitig nimmt die Varianz $\text{Var}(V)$ von V zu und damit auch der Gesamtfehler der Varianz.

Bei asymmetrischer Anordnung der Klassen ist die Kovarianz immer (vgl. Beispiel VII)

$$\text{Cov}[X; E(V/X)] > 0 .$$

Fazit:

Für symmetrische Verteilungen der Zufallsvariablen Z kann bei symmetrischer Anordnung der Klassen sowohl die Varianz $\text{Var}(X)$ von X kleiner als auch größer sein als die Varianz $\text{Var}(Z)$ von Z . Sobald die Symmetrieachse der Dichte $f(z)$ von Z durch die Anordnung der Klassen nicht beachtet wird (die Klassen werden asymmetrisch angeordnet), ist der Gesamtfehler der Varianz immer positiv: Die Varianz $\text{Var}(X)$ von X , die durch die klassifizierten Daten geschätzt wird, ist kleiner als die gesuchte Varianz $\text{Var}(Z)$ von Z .

Um die quantitativen Ergebnisse zu bewerten, wurden in Tabelle (3) die betragsmäßigen relativen Gesamtfehler (Betrag des Gesamtfehlers dividiert durch den wahren Parameterwert der Zufallsvariablen Z) zusammengestellt. Die ausgewiesenen Zahlen sprechen für sich (vgl. Tabelle (3) im Anhang).

Zusammenfassend ist zu sagen :

So wie hier bei exponentialverteilter und bei Laplace-verteilter Zufallsvariablen Z der Fehler der Klasseneinteilung untersucht und quantifiziert wurde, so kann man dies für jede andere beliebige Verteilung von Z bei jeder gegebenen Klasseneinteilung tun.

IV. ZUSAMMENFASSUNG UND KRITISCHE WÜRDIGUNG

Ziel der vorliegenden Arbeit war es, die Einflüsse der Klasseneinteilung bei Stichproben auf die Parameterschätzung zu untersuchen und zu quantifizieren.

Im ersten Teil wurde gezeigt, daß eine Klasseneinteilung immer einen Informationsverlust nach sich zieht. Weiterhin: Gerade bei ökonomischen Untersuchungen, die auf Befragungen (Befragungen nach dem Einkommen, dem Vermögen oder der Konsumgewohnheiten) angewiesen sind, werden die Daten oft schon in Gruppen (Klassen) erhoben. Das heißt, die Klasseneinteilung ist a priori gegeben. Damit mußte die Frage geklärt werden: Wie wirkt sich diese Gruppenbildung auf die Parameterschätzung aus?

Bei statistischen Analysen wird nun meist unterstellt, daß die erhobenen Daten sich als Realisationen einer Zufallsvariablen Z auffassen lassen. Ist die Klasseneinteilung a priori gegeben, kann man die Realisationen von Z nicht beobachten, sondern nur die Klassen bzw. ihre Mitten. Damit stand das erste Ergebnis dieser Arbeit fest: Die Gruppenbildung transformiert eine stetig verteilte Zufallsvariable Z in eine diskret verteilte zufällige Veränderliche X .

Die klassifizierte Stichprobe besteht also aus Realisationen der diskret verteilten Zufallsvariablen X . Durch sie werden dann die Parameter der Verteilungsfunktion $F(x)$ von X geschätzt und

nicht die gesuchten Parameter der Verteilungsfunktion $F(z)$ von Z . Über den Zusammenhang der Zufallsvariablen X und Z ließ sich dann eine Beziehung zwischen dem Erwartungswert $E(X)$ von X und dem Erwartungswert $E(Z)$ von Z sowie zwischen der Varianz $\text{Var}(X)$ von X und der Varianz $\text{Var}(Z)$ von Z ableiten. Es ergab sich

$$E(Z) = E(X) + E[E(V/X)]$$

und

$$\text{Var}(Z) = \text{Var}(X) + 2\text{Cov}[X; E(V/X)] + \text{Var}(V).$$

Die Größe von $E[E(V/X)]$ konnte man als Gesamtfehler des Erwartungswertes interpretieren, die von $2\text{Cov}[X; E(V/X)] + \text{Var}(V)$ als Gesamtfehler der Varianz auffassen. Mit Hilfe dieser Ergebnisse ist man in der Lage, für jede beliebige Verteilung $F(z)$ von Z und jede a priori vorgegebene Klasseneinteilung, den hieraus entstehenden Fehler zu analysieren und zu quantifizieren. Somit schließt sich der Kreis. In Teil III zeigte ich dieses am Beispiel der Exponential- und der Laplace-Verteilung.

Der Leser mag sich nun fragen: Wenn man unbekannte Parameter durch gruppierte Daten schätzen will, wie kommt man dann zu der entsprechenden Form der

Verteilungsfunktion $F(z)$ von Z , damit die entsprechenden Fehlergrößen bei der Schätzung mitberücksichtigt werden können? Man kann sich zwei Möglichkeiten vorstellen. Entweder ist die Verteilungsform von Z aus Erfahrung bekannt. Oder man verschafft sich die notwendige Vorinformation durch ein "zweistufiges-Verfahren". Bei diesem wird in der ersten Stufe durch eine kleine Stichprobe und genaue Daten die Verteilungsform von Z geschätzt. In einer zweiten lassen sich dann durch gruppierte Daten und die entsprechenden Gesamfehlergrößen die Parameter von Z schätzen. Die weitere Ausgestaltung eines solchen "zweistufigen Verfahrens" muß nachfolgenden Forschungen vorbehalten bleiben.

Interessant wäre natürlich, die Fehler der Klasseneinteilung bei mehrdimensionalen Problemen zu bestimmen, zum Beispiel bei der Regressionsanalyse. In der Arbeit sollte jedoch die grundsätzliche Problematik der Klasseneinteilung herausgearbeitet werden. Eine analoge Vorgehensweise würde auch für mehrdimensionale Fragestellungen zu ganz ähnlichen Ergebnissen führen, also nichts grundlegend Neues bringen.

Zusammenfassend läßt sich sagen: Wenn man Vorinformationen über die Verteilung der Zufallsvariablen Z zur Verfügung hat, kann man mit jeder Art von klassifizierten Daten zu statistisch gesicherten Ergebnissen gelangen. Das heißt: In diesem Fall lassen sich die Klassen allein nach von außen vorgegebenen Einflüssen bilden. Dies können beispielsweise der Abbau von psychologischen Hemmnissen bei den Befragten und (oder) die Verringerung des Arbeitsaufwands der Datenerfassung sein.

A N H A N G

Beispiel I

Exponentialverteilung mit $\lambda = 1$

Anzahl der Klassen: 13

Klassenbreite $2c = 1$

Rechte Randklasse offen

z'_j	z'_{j+1}	x_j	p_j	$E(Z/X=x_j)$	$E(V/X=x_j)$
0	1	0,5	0,632121	0,41802	-0,08198
1	2	1,5	0,232544	1,41802	-0,08198
2	3	2,5	0,085548	2,41802	-0,08198
3	4	3,5	0,031471	3,41802	-0,08198
4	5	4,5	0,011578	4,41802	-0,08198
5	6	5,5	0,004252	5,41802	-0,08198
6	7	6,5	0,001567	6,41802	-0,08198
7	8	7,5	0,000576	7,41802	-0,08198
8	9	8,5	0,000212	8,41802	-0,08198
9	10	9,5	0,000078	9,41802	-0,08198
10	11	10,5	0,000029	10,41802	-0,08198
11	12	11,5	0,000011	11,41802	-0,08198
12	∞	12,5	0,000006	13,00000	0,5

Auswertung: Beispiel I

$$E(Z) = 1 \qquad \text{Var}(Z) = 1$$

$$E(X) = 1,081975 \qquad \text{Var}(X) = 0,920620$$

$$E[E(Z/X)] = 1 \qquad \text{Var}[E(Z/X)] = 0,920702$$

$$E[E(V/X)] = -0,081977 \qquad \text{Var}[E(V/X)] = 0,000002$$

$$E[\text{Var}(Z/X)] = 0,079298 = E[\text{Var}(V/X)]$$

$$\text{Var}(V) = 0,0793$$

$$\text{Cov}[X; E(V/X)] = 0,000039$$

Gesamtfehler des Erwartungswertes:

$$\underline{E[E(V/X)] = -0,081977}$$

Gesamtfehler der Varianz:

$$\underline{2\text{Cov}[X; E(V/X)] + \text{Var}(V) = 0,079378}$$

Beispiel II

Exponentialverteilung mit $\lambda = 1$

Anzahl der Klassen: 8

Klassenbreite: $2c = 1$

Rechte Randklasse offen

z'_j	z'_{j+1}	x_j	p_j	$E(Z/X=x_j)$	$E(V/X=x_j)$
0	1	0,5	0,632121	0,41802	-0,08198
1	2	1,5	0,232544	1,41802	-0,08198
2	3	2,5	0,085548	2,41802	-0,08198
3	4	3,5	0,031471	3,41802	-0,08198
4	5	4,5	0,011578	4,41802	-0,08198
5	6	5,5	0,004259	5,41802	-0,08198
6	7	6,5	0,001567	6,41802	-0,08198
7	∞	7,5	0,000912	8	0,5

Auswertung: Beispiel II

$$E(Z) = 1 \quad \text{Var}(Z) = 1$$

$$E(X) = 1,081446 \quad \text{Var}(X) = 0,912719$$

$$E[E(Z/X)] = 1 \quad \text{Var}[E(Z/X)] = 0,919837$$

$$E[E(V/X)] = -0,081449 \quad \text{Var}[E(V/X)] = 0,000309$$

$$E[\text{Var}(Z/X)] = 0,08016 = E[\text{Var}(V/X)]$$

$$\text{Var}(V) = 0,080469$$

$$\text{Cov}[X; E(V/X)] = 0,0012690$$

Gesamtfehler des Erwartungswertes:

$$E[E(V/X)] = -0,081449$$

Gesamtfehler der Varianz:

$$2 \text{ Cov}[X; E(V/X)] + \text{Var}(V) = 0,083007$$

Beispiel III

Exponentialverteilung mit $\lambda = 1$

Anzahl der Klassen: 5

Klassenbreite, $2c = 1$

Rechte Randklasse offen

z_j	z'_{j+1}	x_j	p_j	$E(Z/X=x_j)$	$E(V/X=x_j)$
0	1	0,5	0,632121	0,41802	-0,08198
1	2	1,5	0,232544	1,41802	-0,08198
2	3	2,5	0,085548	2,41802	-0,08198
3	4	3,5	0,031471	3,41802	-0,08198
4	∞	4,5	0,018316	5	0,5

Auswertung: Beispiel III

$$E(Z) = 1 \quad \text{Var}(Z) = 1$$

$$E(X) = 1,0713377 \quad \text{Var}(X) = 0,824628$$

$$E[E(Z/X)] = 1 \quad \text{Var}[E(Z/X)] = 0,903814$$

$$E[E(V/X)] = -0,071320 \quad \text{Var}[E(V/X)] = 0,006090$$

$$E[\text{Var}(Z/X)] = 0,096189 = E[\text{Var}(V/X)]$$

$$\text{Var}(V) = 0,102279$$

$$\text{Cov}[X; E(V/X)] = 0,036547$$

Gesamtfehler des Erwartungswertes:

$$E[E(V/X)] = -0,071320$$

Gesamtfehler der Varianz:

$$2\text{Cov}[X; E(V/X)] + \text{Var}(V) = 0,175373$$

Beispiel IV

Exponentialverteilung mit $\lambda = 1$ 、

Anzahl der Klassen: 3

Klassenbreite: $2c = 1$

Rechte Randklasse offen

z_j	z'_{j+1}	x_j	p_j	$E(Z/X=x_j)$	$E(V/X=x_j)$
0	1	0,5	0,632121	0,41802	-0,08198
1	2	1,5	0,232544	1,41802	-0,08198
2	∞	2,5	0,135335	3	0,5

Auswertung: Beispiel IV

$$E(Z) = 1 \qquad \text{Var}(Z) = 1$$

$$E(X) = 1,003214 \qquad \text{Var}(X) = 0,520656$$

$$E[E(Z/X)] = 1,00000 \qquad \text{Var}[E(Z/X)] = 0,796075$$

$$E[E(V/X)] = -0,003186 \qquad \text{Var}[E(V/X)] = 0,520656$$

$$E[\text{Var}(Z/X)] = 0,203925 = E[\text{Var}(V/X)]$$

$$\text{Var}(V) = 0,243555$$

$$\text{Cov}[X; E(V/X)] = 0,117858$$

Gesamtfehler des Erwartungswertes:

$$E[E(V/X)] = -0,003186$$

Gesamtfehler der Varianz:

$$2\text{Cov}[X; E(V/X)] + \text{Var}(V) = 0,479271$$

Beispiel V

Laplace-Verteilung mit $E(Z) = 0$

und $\text{Var}(Z) = 32$ ($\lambda = 4$)

Anzahl der Klassen: 14

Klassenbreite: $2c = 4$

Rechte und linke Randklasse offen

z'_j	z'_{j+1}	x_j	p_j	$E(Z/X=x_j)$	$E(V/X=x_j)$
∞	-24	-26	0,001239	-28	-2
-24	-20	-22	0,002130	-21,672093	+0,327907
-20	-16	-18	0,005789	-17,672093	+0,327907
-16	-12	-14	0,015736	-13,672093	+0,327907
-12	- 8	-10	0,042774	- 9,672093	+0,327907
- 8	- 4	- 6	0,116272	- 5,672093	+0,327907
- 4	0	- 2	0,316060	- 1,672093	+0,327907
0	4	2	0,316060	+ 1,672093	-0,327907
4	8	6	0,116272	+ 5,672093	-0,327907
8	12	10	0,042774	+ 9,672093	-0,327907
12	16	14	0,015736	+13,672093	-0,327907
16	20	18	0,005789	+17,672093	-0,327907
20	24	22	0,002130	+21,672093	-0,327907
24	∞	26	0,001239	+28	+2

Auswertung: Beispiel V

$$E(Z) = 0 \quad , \quad \text{Var}(Z) = 32$$

$$E(X) = 0 \quad \text{Var}(X) = 33,111616$$

$$E[E(Z/X)] = 0 \quad \text{Var}[E(Z/X)] = 30,694227$$

$$E[E(V/X)] = 0 \quad \text{Var}[E(V/X)] = 0,117169$$

$$E[\text{Var}(Z/X)] = 1,305774 = E[\text{Var}(V/X)]$$

$$\text{Var}(V) = 1,422943$$

$$\text{Cov}[X; E(V/X)] = -1,316341$$

Gesamtfehler des Erwartungswertes:

$$E[E(V/X)] = 0$$

Gesamtfehler der Varianz:

$$2 \text{ Cov}[X; E(V/X)] + \text{Var}(V) = -1,209739$$

Beispiel VI

Laplace-Verteilung mit $E(Z) = 0$

und $\text{Var}(Z) = 32$ ($\lambda = 4$)

Anzahl der Klassen: 8

Klassenbreite: $2c = 4$

Rechte und linke Randklasse offen

z_j	z'_{j+1}	x_j	p_j	$E(Z/V=x_j)$	$E(V/X=x_j)$
$-\infty$	-12	-14	0,024894	-16	-2
-12	-8	-10	0,042774	-9,672093	+0,327907
-8	-4	-6	0,116272	-5,672093	+0,327907
-4	0	-2	0,316060	-1,672093	+0,327907
0	4	2	0,316060	+1,672093	-0,327907
4	8	6	0,116272	+5,672093	-0,327907
8	12	10	0,042774	+9,672093	-0,327907
12	∞	14	0,024894	+16	+2

Auswertung: Beispiel IV

$$E(Z) = 0 \quad , \quad \text{Var}(Z) = 32$$

$$E(X) = 0 \quad \text{Var}(X) = 29,213312$$

$$E[E(Z/X)] = 0 \quad \text{Var}[E(Z/X)] = 29,997586$$

$$E[E(V/X)] = 0 \quad \text{Var}[E(V/X)] = 0,301322$$

$$E[\text{Var}(Z/X)] = 2,002414 = E[\text{Var}(V/X)]$$

$$\text{Var}(V) = 2,303736$$

$$\text{Cov}[X; E(V/X)] = 0,241476$$

Gesamtfehler des Erwartungswertes:

$$E[E(V/X)] = 0$$

Gesamtfehler der Varianz:

$$2 \text{ Cov}[X; E(V/X)] + \text{Var}(V) = 2,786688$$

Beispiel VII

Laplace-Verteilung mit $E(Z) = 0$

und $\text{Var}(Z) = 32$ ($\lambda = 4$)

Anzahl der Klassen: 5

(asymmetrische Anordnung der Klassen)

Klassenbreite $2c = 4$

Rechte und linke Randklassen offen

z'_j	z'_{j+1}	x_j	p_j	$E(Z/X=x_j)$	$E(V/X=x_j)$
- ∞	0	- 2	0,5	-4	-2
0	4	2	0,316060	1,672093	-0,327907
4	8	6	0,116272	5,672093	-0,327907
8	12	10	0,042774	9,672093	-0,327907
12	∞	14	0,024894	16	2

Auswertung: Beispiel VII

$$E(Z) = 0 \quad \text{Var}(Z) = 32$$

$$E(X) = 1,106008 \quad \text{Var}(X) = 15,383402$$

$$E[E(Z/X)] = 0 \quad \text{Var}[E(Z/X)] = 22,998740$$

$$E[E(V/X)] = -1,106002 \quad \text{Var}[E(V/X)] = 0,927419$$

$$E[\text{Var}(Z/V)] = 9,00126 = E[\text{Var}(V/X)]$$

$$\text{Var}(V) = 9,928679$$

$$\text{Cov}[X; E(V/X)] = 3,343985$$

Gesamtfehler des Erwartungswertes:

$$E[E(V/X)] = -1,106002$$

Gesamtfehler der Varianz:

$$2\text{Cov}[X; E(V/X)] + \text{Var}(V) = 16,616649$$

Tabellen (1) und (2)

Tabelle (1)

Die Gesamtfehler - Exponentialverteilung

Anzahl d. Klassen	$E[E(V/X)]$	$Cov[X; E(V/X)]$	$Var(V)$	Gesamt- fehler d. Varianz
13	-0,081977	0,000039	0,079300	0,079378
8	-0,081449	0,001269	0,080469	0,083007
5	-0,071320	0,036547	0,102779	0,175373
3	-0,003186	0,117858	0,243555	0,479271

Tabelle (2)

Die Gesamtfehler - Laplace-Verteilung

Anzahl d. Klassen	$E[E(V/X)]$	$Cov[X; E(V/X)]$	$Var(V)$	Gesamt- fehler d. Varianz
14	0	-1,316341	1,422943	-1,209739
8	0	0,241476	2,303736	2,786688
5 ⁽¹⁾	-1,106002	3,343985	9,928679	16,616649

(1) asymmetrische Anordnung der Klassen

Tabelle (3)

Betragsmäßige relative Fehler⁽¹⁾

Anzahl d. Klassen	von $E(Z)$	von $Var(Z)$
----------------------	------------	--------------

1) Exponentialverteilung

13	8,20%	7,94%
8	8,14%	8,30%
5	7,13%	17,54%
3	0,32%	47,92%

2) Laplace-Verteilung

14	-	3,78%
8	-	8,71%
5(2)	-	51,93%

(1) Betragsmäßiger relativer Fehler=

$$\frac{\text{Betrag des Gesamtfehlers}}{\text{wahren Parameterwert von } Z} \cdot 100$$

(2) asymmetrische Anordnung der Klassen

LEBENS LAUF

Persönliche Daten

Geburtsdatum	13.06.1951
Geburtsort	Bensheim/Bergstraße
Staatsangehörigkeit	deutsch
Familienstand	ledig

Bildungsgang

Grundschule	1958-1963	Schillerschule Groß-Gerau
Gymnasium	1963-1972	Prälat-Diehl-Schule Groß-Gerau
	1972	Abitur
Hochschule	1972-1978	Johannes Gutenberg-Universität Mainz
	WS 72/73	Studienfach Chemie
	ab SS 73	Studienfach Volkswirtschaftslehre
	Mai 78	Diplomprüfung für Volkswirte

Berufspraxis

01.10.1978-28.02.1979	Trainee-Programm im Versicherungs- und Bankwesen
ab 01.03.1979	wissenschaftlicher Mitarbeiter am Fachbereich Rechts- und Wirtschaftswissenschaften der Johannes Gutenberg-Universität Mainz
15.02.1984	mündliche Doktorprüfung

Mainz, den 21.02.1984

Fz - Ferdinand

difo
druck

laubanger 15
8600 bamberg
tel.: 0951/6 4321